PREDICTING VAGINAL BIRTH AFTER CESAREAN DELIVERY SUCCESS: IS IT POSSIBLE?

by

Jason Naoki Hashima

A THESIS

Presented to the Department of Public Health and Preventive Medicine

and the Oregon Health & Science University

School of Medicine

in partial fulfillment of

the requirements for the degree of

Master of Public Health

April 2003

School of Medicine
Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the Master of Public Health thesis of

Jason Naoki Hashima

has been approved

[redacted]
Chairperson – Jeanne Marie Guise, MD, MPH

[redacted]
Member – Cynthia Morris, PhD, MPH

[redacted]
Member – Jodi Lapidus, PhD

Associate Dean for Graduate Studies

STRUCTURED ABSTRACT

Background. The question of what route of delivery should be taken in women with a prior cesarean delivery (CD) remains unanswered nearly a century after Edwin B. Cragin stated the dictum of "once a cesarean, always a cesarean". Among many issues, the return to rising CD rates during the last five years has prompted the need of a method to select individuals suitable for a trial of labor (TOL) following a prior CD. This study aims to determine what factors predict the route of delivery for patients with a prior cesarean delivery who undergo a trial of labor.

Methods. A meta-analysis was performed using classical random effects modeling to identify factors associated with a successful TOL (i.e. vaginal delivery) following prior CD. Heterogeneity was explored using the Breslow Day test and subgroup analyses.

Data Sources. Studies were identified in MEDLINE® and HealthSTAR® databases from 1980 to March 2002. Reviewing reference lists identified additional studies.

Study Selection. Dual review selection of the articles was based on a pre-determined set of inclusion criteria.

Data Extraction. Of the 2,945 originally identified titles and abstracts, 249 were pulled for further review. 100 met all of the inclusion criteria.

Results. Summary odds ratios were calculated for 12 factors providing categorical data and summary mean comparisons for eight factors providing continuous data. Factors found to be significantly associated with an increased likelihood of a successful TOL (i.e. vaginal delivery) include parity greater than one, only one prior CD, previous VD (especially after prior CD), non-recurrent prior CD indication (including breech or fetal distress), spontaneous onset of labor, no use of oxytocin (including augmentation) and infant weight less than 4000g. The comparison of summary mean estimates for those with a successful TOL (i.e. vaginal birth) and those with a failed TOL demonstrated that those with a vaginal delivery were younger and weighed less, and had an infant that weighed less. The exploration of study heterogeneity revealed that study design, quality, and size only partially explained the observed variability.

Conclusions. This study has demonstrated that several factors are associated with TOL outcome. The use of more specific categorization of each factor and the recognition and adjustment of confounding will provide increasing validity of future studies. Study heterogeneity appears to be influenced by a combination of different study characteristics and potentially other unexplored factors.

TABLE OF CONTENTS

TABLES

FIGURES

ACKNOWLEDGEMENTS

There are many people I wish to thank for helping me get to where I am today. Every step of my life, I have been fortunate to be around and learn from very special individuals. It is from their support and guidance that I have come as far as I have. The achievement of obtaining the Master of Public Health is one that I will be proud of for the rest of my life. The best way to describe the work required to finish this degree in conjunction with my medical training is by quoting one of Jodi Lapidus' students in that the process was "*grueling, yet strangely rewarding*". Below I have listed some of those individuals who have contributed to my MPH education. To them I say "thank you".

Jeanne Marie Guise

Cindy Morris

Jodi Lapidus

Benjamin Chan

Peggy Nygren

Patricia Osterweil

Karen Eden

Patty Davies

Marian McDonaugh

Thomas Becker

John Stull

OHSU Evidence based Practice Center

The Agency for Healthcare Research and Quality

MPH faculty and staff

Lisa Fitzgerald

Family and Friends

Background/Significance:

The first reported transperitoneal, retrovesical, transverse incision for the purposes of cesarean delivery (CD) was performed by Ferdinand Adolph Kehrer of Heidelberg, Germany in 1881, which according to many historians, marked the birth of the era of medically determined abdominal delivery.[1] Although the high associated morbidity and mortality prevented the widespread use of such procedures prior to the time, advancements in medicine during the late 1800's such as the advent of obstetric anesthesia, adoption of the antiseptic techniques advocated by Joseph Lister, and the suturing of the uterine incision as proposed by Sanger collectively increased the safety and enthusiasm for the procedure.[2]

On May 12, 1916, Edwin B. Cragin stated near the end of his address to the Eastern Medical Society of the City of New York that, "One thing must always be borne in mind, that no matter how carefully a uterine incision is sutured, we can never be certain that the cicatrized uterine wall will stand a subsequent pregnancy and labor without rupture. The means that the usual rule is, *once a cesarean, always a cesarean.*"[3] Although Cragin did not deny the possibility of a vaginal birth after cesarean delivery (VBAC), his intent was simply to urge physicians to avoid unnecessary cesarean operations and to use it only as a last resort. While the dictum of "once a cesarean, always a cesarean" represented the majority opinion of the time, several clinicians and investigators questioned such a philosophy, including Kerr who in 1926 introduced the low-transverse uterine incision for which he claimed that once healed, such an incision would permit a safe labor in subsequent pregnancies. The decades to follow were filled with controversy. As with many other debates in medicine, the question of what is the

appropriate birth route in the subsequent pregnancy of those with a prior CD remains largely unanswered.

In 1970, the reported CD rate of 5.4 per 100 live births marked the beginning of a period of dramatic increase in the rate. By 1980, in comparison to other countries of the world, the United States had one of the highest CD rates (16.9 per 100 live births), making it the tenth most common surgical procedure in the United States. Compared to other countries of the world, the United States also had one of the lowest vaginal birth after cesarean (VBAC) rates (3.0 per 100 live births).[4] Although VBAC was common in Europe even before 1950, it remained relatively uncommon in the United States. The increase in CDs was speculated to result from the fear of malpractice, increased use of electronic fetal monitoring, lack of healthcare provider training and experience in breech delivery, and decreased use of operative vaginal deliveries.[5] Of the proposed reasons, the most common indication for CD was elective repeat cesarean delivery (ERCD, i.e. the patient did not require a CD, but instead elected to have one), which by 1984 was found in one study to comprise 36% of all cesarean operations.[6]

In response to rising CD rates and consumer requests for CDs, the U.S. National Institute of Child Health and Human Development convened the Conference on Childbirth in 1980. From their work, they concluded that 25-30% of the observed overall increase in CD rate was contributed by ERCD and that VBAC was an appropriate manner by which to decrease the rising cesarean rates.[7] Soon after, the American College of Obstetricians and Gynecologists (ACOG) issued clinical management guidelines for the use of VBAC.[8] Since that time, the concept of "once a cesarean, always a cesarean" began to slowly change as improvements in obstetric care made a trial of labor safer for

both the mother and infant. In fact, during the 1980s, numerous articles were published on VBAC, including large multicenter studies,[9,10] which confirmed the relative safety of VBAC. By 1993, the national VBAC rate climbed to 25 per 100 live births[11] and the number of VBACs exceeded 100,000 per year. A far cry from the mere 100 cases reported by the Margaret Hague Maternity Hospital in New Jersey between 1931 and 1950.[12]

During the early 1990s, several studies were able to demonstrate that a trial of labor (TOL) was successful in 60-80% of VBAC attempts, while showing no significant effect in the maternal or neonatal morbidity and mortality.[9,10,13] However in 1991, two reports[14,15] raised serious doubt over the safety of VBAC. These case reports demonstrated a uterine rupture rate (less than 1 per 100 TOL attempts) similar to those reported previously. However the concern arose from the high percentage of catastrophic outcomes, including perinatal deaths and long-term neurologic deficits. Although an accompanying editorial by Pitkin[16] stated that these uncontrolled observational studies in no way negated the conclusions from dozens of large cohort studies that indicated a low complication rate of VBAC, many physicians and patients began to wonder if they should return to the philosophy of "once a cesarean, always a cesarean". The concern regarding VBAC was further heightened in 1996 after several major U.S. newspapers printed the headlines "VBAC twice as risky as repeat cesarean," following the release of a large population-based Canadian study by McMahon.[12] This study demonstrated that major maternal complications, such as uterine rupture, operative injury, and a need for hysterectomy were nearly twice as likely among women undergoing a TOL, than among women undergoing an elective CD.[17] Despite the possibility that a more relaxed

approach to TOL attempts (e.g. allowing breech deliveries, use of more aggressive induction and augmentation, etc.) could have contributed to the increase in complication rates, the confidence in VBAC began to diminish.

At the same time subgroup analyses suggested that the main source of TOL maternal and neonatal morbidity, occurred for women who failed their TOL[17,18]. When all TOLs (i.e. successful and failed attempts) were considered, there was no statistical difference in the complication rates between TOL and ERCD. In a meta-analysis by Rosen, failed TOL was associated with an increased risk of uterine dehiscence or rupture compared to ERCD (odds ratio (OR) 2.8; 95% confidence interval (CI) 1.4-5.4). The risk was not shown when all of those attempting a TOL were compared to those undergoing ERCD (OR 0.8; 95% CI 0.6-1.2).[18]

Considering the finding, many researchers attempted to identify maternal and obstetric factors associated with a failed TOL in order to minimize the number of major maternal and neonatal complications. More than one hundred published studies look specifically at those risk factors and their influence on the route of delivery following prior CD. Several general clinical reviews of the VBAC literature have been published of the factors that influence the outcome of a TOL.[19,20,21] Unlike the reviews by Lavin[19] and Weinstein[20], McMahon[21] rated the included studies according to the levels of evidence proposed by the United States Preventive Services Task Force. From his review, McMahon concluded that the existing literature was in fact insufficient (i.e. lack of quality data), thus making any conclusions drawn from them invalid.

Table 1. Individual factors included in past reviews

	Lavin et al (1982)	Weinstein et al (1996)	McMahon (1998)
Years of literature review	1950-1980	NR	NR
Factors			
Number of prior CD	X	X	
Prior CD indication	X	X	X
Previous Vaginal Delivery (VD)	X		X
Oxytocin use	X	X	
Prostaglandin use		X	
Epidural/Anesthesia	X	X	
Breech	X	X	X
Twins	X	X	X
Macrosomia			X
X-Ray Pelvimetry (XRP)	X		

NR = Not Reported; X = factor reviewed

These reviews were limited by the lack of a systematic approach and the use of "vote counting" for purposes of analysis (i.e. summing crude data to calculate pooled estimates). In 1990, Rosen attempted to take the review process one step further by providing summary estimates for the factors that influence the route of delivery.[22] By reviewing studies published from 1982-1989, Rosen found that the following five factors were significantly associated with successful TOL: non-recurrent indication for prior CD, breech indication for prior CD, previous vaginal delivery (VD), only one prior CD, an no oxytocin use. Rosen was able to produce more valid estimates compared to previous reviews by statistically accounting for within-study variation and by adjusting for each study as a confounder. However, Rosen's analysis was not without fault. The first limitation involved using the Mantel-Haenszel method (i.e. fixed effects modeling) for pooling the odds ratios and calculating confidence interval. While it may be considered a

legitimate approach for combining data and adjusting for confounding, the validity of the process depends on the assumption that all of the effects in these studies were relatively homogeneous. Because Rosen failed to provide any information regarding the evaluation of heterogeneity between studies (e.g. Breslow-Day statistic) one is left to question the validity of the estimates. Another limitation of Rosen's study, as well as the majority of the other published studies in the VBAC literature, is the lack of accounting for confounding in the observed associations. Although he states that the analysis was limited to those studies with similar comparison groups, the degree of similarity is not mentioned. The potential for estimate distortion by confounding, requires the use of caution in the interpretation of his results.

While many investigators concentrated on individual factors and their influence on the route of delivery, others have attempted to look at the combined predictive ability of several maternal and obstetric factors in the form of a scoring system.[23,24] While two of these scoring systems have been validated in different patient populations, the lack of their widespread use could indicate the lack of confidence in their findings. In his own investigation, Pickhardt[28] concluded that he was not able to identify a single criterion or an optimal cluster of factors that predicted successful TOL and that no synthesized equation was able to achieve a greater than 75% predictive value of outcome with an acceptable sensitivity and specificity.

The ability to accurately and reliably predict vaginal delivery in those with a prior CD would be truly invaluable. The approximate figure of four million births occurring in the U.S. each year, translates to more than 10,000 babies born each day. One in five infants (i.e. 22.9% of all births) are born by CD,[29] which includes approximately one

third that are by elective repeat cesarean births.[30] In the end, nearly 250,000 women per year undergo the elective procedure. Although studies have been able to demonstrate that 60-80% of those who undergo a TOL have a vaginal delivery, some women and healthcare providers may feel that the 20-40% probability of requiring an emergent CD along with the increased risk of morbidity and mortality outweighs the desire for a vaginal delivery. By being able to accurately and reliably predict the probability of vaginal delivery, one could potentially reduce the number of CDs performed each year by educating and encouraging those women with a high probability of success to undergo a TOL. Based on a hypothetical model of a 30-yr old patient and a threshold of $50,000 per quality-adjusted life years to define cost-effectiveness, Chung et al demonstrated a positive cost-benefit ratio between a TOL and ERCD depending on the probability of VD.[31] For example, if the probability of VD was less than or equal to 74%, then ERCD was more cost-effective; and if the probability of VD was greater than 74%, then TOL was more cost-effective. Lastly, the ability to accurately and reliably predict the probability of VD is important because of the changes in the 1999 ACOG guidelines for VBAC, which called for the reduction in the allowable time between the onset of prolonged deceleration and delivery from the previously established 30 minutes.[30] This was based on an observation that significant neonatal morbidity occurred when $\geq$ 18 minutes elapsed between the onset of prolonged deceleration and delivery.[32] The ACOG guidelines created doubt for many health care providers as to whether or not their facilities and staff could comply with such recommendations. The fear of complications such as uterine rupture and not being able to provide timely care for the patient in such a situation has led many to question the utility and feasibility of VBAC.

Objective:

The objective of this study is to perform a systematic review of the literature to provide a better understanding of the existing VBAC literature regarding predictive factors, and in turn provide health care providers, patients, and researchers with a clearer and stronger foundation upon which to make informed decisions.

Study Design and Methods:

1. *Literature Search and Selection of Articles*:

With the assistance of a medical librarian, searches were conducted in MEDLINE (1966 to 2002) and HealthSTAR (1975-2002) to identify relevant studies using specific search terms such as "vaginal birth after cesarean", "trial of labor", and "trial of scar" (see Appendix A). Titles and abstracts were captured and placed into an electronic database (Endnote®).

To reduce observer bias, two investigators performed an independent review of a random subset of 200 titles and abstracts, during which specific inclusion criteria were examined. Studies were considered for inclusion if they: 1) included women with a prior CD, 2) included an analysis of a factor and the outcome of successful or failed TOL (i.e. data useful for the purposes of the report), 3) included any study design besides a case report or case series (less than 10 subjects), 4) were not a letter or an editorial, 5) were available in the English language, 6) were published in a foreign language and provided information inconsistent with those studies published in English, 7) were conducted after 1980, 8) were from a developed country, 9) did not include breech delivery, and 10) did not focus on patients with vertical or classical incisions. The decision to limit the search

to those studies that were conducted after 1980 was based on the fact that in 1980, the U.S. National Institute of Child Health and Human Development Conference on Childbirth publicly stated that a TOL was an acceptable delivery option in those with a previous cesarean scar. As the announcement could have changed the obstetric practice, literature following such a decision was considered pertinent to this study. Only data from developed countries would be used in the analysis because these countries would most likely have similar obstetric conditions and practices as to those in the Untied States. The CIA World Factbook 2001 and the International Monetary Fund list listed 39 countries as being developed (see Appendix B). Due to the high costs of translation, studies published in a foreign language were only considered for inclusion if they provided information that was inconsistent with the findings of studies published in English. This method of selection would also minimize the risk of committing the "tower of Babel" bias[34], by excluding sound studies published in another language. Lastly, the decision to not include those studies that focused on breech TOL attempts was based on the recent ACOG Practice Guideline,[35] which recommended against vaginal breech deliveries. The statement was made after a recent international multicenter randomized clinical trial of breech delivery showed a significant increase in neonatal morbidity and mortality for those attempting a TOL compared to those with an elective CD.[36]

A reliability analysis using a kappa statistic was performed upon the completion of the independent reviews. Disagreements over study inclusion (i.e. yes or no) were discussed between the two reviewers until a consensus could be reached regarding the inclusion or exclusion of a particular study. Once an appropriate level of reliability was achieved (i.e. kappa statistic greater than 0.80), the primary investigator reviewed the

remaining titles and abstracts using the agreed upon inclusion and exclusion parameters. Although no direct author contact was required in this study, the effort to include all available studies involved the evaluation of the reference lists for each of the articles identified in the initial review of titles and abstracts for any additional studies, which might be pertinent to the investigation. The combination of the original full-text articles and those articles identified through the reference lists were then once again re-evaluated according to the previously mentioned inclusion criteria. Those studies, which met all inclusion criteria, were then subject to data extraction.

2. *Quality*:

Studies fulfilling the inclusion criteria at the full-text article level were subsequently evaluated for quality. To reduce observation/investigator bias, two investigators performed a dual review for quality assessment of a subset of 20 full-text articles. The two investigators discussed any discrepancies in ratings, and resolved these differences by consensus. Once the quality rating process was established, the primary investigator was responsible for evaluation of the remaining full-text articles.

The quality criteria used in this study were adapted from those established by the United States Preventive Services Task Force (USPSTF),[148] as well as those developed by the National Health Services (NHS) Center for Reviews and Dissemination, based at the University of York in England.[149] The applied quality ratings were unique to each study design and are as follows:

RCTs. Based on the USPSTF and York guidelines, eleven criteria were evaluated for each RCT study. These criteria included: random assignment to the specific treatment arm or failing to report randomization methods, allocation concealment, comparability of

groups at baseline, explicit eligibility criteria, blinded outcomes assessment, blinding of providers and patients, use of intention-to-treat analysis, maintenance of comparable groups, and the reporting of crossover or attrition and loss to follow up. Four of these eleven criteria were felt to be the most important and were subsequently used as the basis of our quality rating: random assignment to the specific treatment arm or reporting of randomization method, comparability of groups at baseline, use of intention-to-treat analysis, and reporting of loss to follow up. A "good" quality study satisfied all four criteria. A "fair" quality had to at least satisfy the criteria of random assignment to the specific treatment group and the use of intention-to treat analysis. Studies were rated as "poor" quality if they failed to have the random assignment or use of intention-to-treat analysis.

Cohort Studies. Based on the USPSTF and York guidelines, eight criteria were evaluated for each cohort study. These criteria included: the assembly of comparable groups, the maintenance of comparable groups, the clear definition of comparison groups with sufficient description of prognostic factor distribution, measures equal and outcomes well defined, the blind assessment of outcome, follow up proportion of at least 80 percent, follow up period long enough for outcome to occur, and the adjustment of potential confounders. Based on the nature of the topic being investigated, it was decided that several of these criteria would not be as useful (e.g. since the outcome of interest was VD or emergent CD, the criteria of blinded outcome assessment was not as heavily weighted). In the end, it was decided that three of these criteria were the most important in determining the quality of each cohort study: comparable groups assembled, sufficient

description of prognostic factor distribution, and the consideration/adjustment of potential confounders. Table 2 provides a detailed explanation of each of these criteria.

Table 2. Explanation of criteria used for quality ratings

Criteria	Definition
Comparable groups	• Groups being evaluated are comparable in terms of their likelihood to undergo a TOL following prior CD • Some studies are listed with (*) to indicate that the overall study includes non-comparable groups (e.g. oxytocin use in those with PCD and those without PCD), however information regarding certain factors was available for PCD group.
Description of the distribution of prognostic factors	• Groups were clearly defined (i.e. inclusion and exclusion criteria) • Study provides information regarding the distribution of prognostic factors (confounders) for each group.
Consider/adjust for important confounders	• The analysis and discussion of the study employs methods to adjust for confounding (e.g. multivariate or regression analysis) in the observed associations.

Table 3 illustrates how the ratings of "good, fair, or poor" were assigned to each study based upon the three criteria. Although some may argue that these criteria are too stringent, resulting in "poor" ratings for many of the studies evaluated, it was felt that such a rating system allows for the most accurate interpretation of the associations reported. The studies rated as "good" or "fair" in some way attempt to control for these confounding factors, thus making the evaluation of a factors' influence on TOL outcome more valid.

Table 3. Study quality scores as derived from criteria

Quality Score	Criteria Requirements
Good	• "YES" to all three criteria.
Fair	• "YES" to comparable groups and adjustment for confounders and "NO" for clear definition of groups.
Poor	• "YES" to comparable groups and "NO" to clear definition of groups and adjustment for confounders. • "NO" to comparable groups.

Case-Control Studies. Nine quality criteria were used: explicit definition of cases, state of cases validated, accurate ascertainment of cases, nonbiased selection of cases and controls, cases and controls comparable with respect to confounders, procedures applied equally, measurement of exposure accurate and applied equally, appropriate attention to confounders, and appropriate statistical analysis used. Four of these criteria were considered to be the most important and were used as the basis for rating quality. A study was rated as being "good" quality if it satisfied all four criteria of explicit definition of cases, nonbiased selection of cases and controls, measurement of exposure accurate and applied equally, and consideration of confounders. A study was given a "fair" rating if the cases were explicitly defined and some attention was paid to confounders. However, if these two criteria were neglected, then the study was given a "poor" rating.

Case-Series Studies. Six criteria were applied to all case-series studies of greater than 10 subjects. These criteria included: the use of a representative sample from a relevant population, the use of explicit inclusion criteria, all the individuals entered at a similar point in their disease/status progression, follow up was long enough for events to occur, outcomes assessed using objective criteria/blinding, and sufficient description of the distribution of prognostic factors. From these it was decided that three would form

the basis of the quality ratings for case series. A "good" quality study accounted for all three criteria of using a representative sample from a relevant population, all the individuals entered at a similar point in their disease/status progression, and sufficient description of the distribution of prognostic factors. A study received a "fair" quality rating if they at least used a representative sample from a relevant population and the individuals all entered at a similar point in their disease/status progression. If none of these criteria were met, then the study received a "poor" rating.

3. *Data Extraction*:

Using a predesigned worksheet in Microsoft Excel®, data was abstracted on the following study characteristics: author, year of publication, journal, study design, study research objective, years of study, country/site of study, study population, study exclusion criteria, number of subjects in the study, number of subjects attempting a trial of labor, percentage of subjects attempting a trial of labor, number of subjects with a VD, and of those who attempted a trial of labor, the percentage of subjects with a VD.

In addition to these study characteristics, pertinent data regarding various individual factors and their possible association with the route of delivery was also abstracted (e.g. the total number of subjects with that factor, the number of subjects with that factor who had a VD, the total number of subjects without that factor, and the number of subjects without that factor who had a VD). Table 4 provides a list of these factors, grouped according to the general categories of: demographic, past obstetric, current obstetric, and non-clinical factors.

In the situation when there were multiple reports from one study population or when study populations overlapped with each other, only the data from the most comprehensive report was utilized for the purposes of analysis.

Table 4. Individual factors which may influence the route of delivery, grouped by general categories

Category	Factors	
Demographic	Age Race	SES
Past Obstetric	Gravidity Parity Previous VD Previous VD order	Previous Cervical Dilation Number of prior CD Prior CD Indications
Current Obstetric	Gestational age Birth weight Estimated fetal weight Multiple gestations Breech/External Cephalic Version Cervical dilation Cervical dilation rate Cervical effacement Station	Bishop score Spontaneous labor Induced labor Augmented labor Oxytocin use (non-specified) Epidural use Labor duration Maternal height Maternal weight Maternal weight gain
Non-Clinical	Insurance status Hospital characteristics	Physician characteristics

4. *Statistical Analysis*:

Summary Estimates

The common belief that the primary purpose of meta-analysis is to combine the results of several studies requires the underlying assumption that the studies to be combined are homogeneous. As it is considered inappropriate to combine the results of several studies if they significantly different from each other (i.e. heterogeneous), several methods for combining data have been developed (e.g. fixed effects models for

homogeneous data and random effects models for heterogeneous data). This study elected to utilize a random effects model, regardless of the homogeneity or heterogeneity of the study estimates. In doing so, the more conservative approach would take into consideration the within and between study heterogeneity, as well as the residual variability inherent in each study.

As described in Table 5, a minimum of three studies was required for calculation of a summary estimate. Although possible, it was felt that the combination of only two studies would not provide any additional information. Summary estimates were also only calculated for those studies that provided data for their individual factors in a similar fashion. Summary mean estimates were calculated for those studies providing continuous data. The calculation of summary mean estimates required that at least three studies provided both the mean and standard deviations for those with a successful and a failed TOL.

Table 5. The requirements for the calculation of summary estimates

1. A minimum of three studies had to be available for the specific factor to be combined.
2. The studies being combined had to provide data in a similar fashion: a. Categorical data: utilize the same group cut points (e.g. birthweight less than 4000 grams versus greater than 4000 grams). b. Continuous data: the mean and standard deviation for both those with a successful and failed TOL were available.

Categorical factors meeting these requirement were visually evaluated using forest plots (i.e. the distribution of odds ratios and confidence intervals of each individual study) created by using the StatsDirect® statistical program. The program was then used to calculate the random effects summary estimates using the Classical approach (i.e.

DerSimonian-Laird pooled odds ratio[37, 38]) and 95% Confidence Intervals for those factors with categorical data. For comparative purposes, the random effects summary estimates were also calculated using the simulation based WINBUGS® statistical program,[39-41] which follows the Bayesian approach. The analysis of the factors presenting continuous data was performed using the WINBUGS® statistical program. By using programming codes (see Appendix C) that required minor editing for each factor (e.g. changing the number of studies to be combined and the input data values), summary mean estimates were calculated for those with a successful TOL and for those with a failed TOL. The calculation not only provided the opportunity for a comparison between those who succeeded at a TOL and those who failed for a specific factor, but the exercise also allowed for the estimation of potential categorical cut-points for future research using these factors.

Publication Bias.

Some have suggested that one could investigate abstract databases or even contact investigators to reduce publication bias. But as these avenues are time consuming and of relatively poor yield, we decided to evaluate publication bias visually using funnel plots and statistically using the Egger proposed regression of normalized effect versus precision. In this regression analysis, one could conclude that a publication bias existed if the intercept significantly differed from zero. Both of these tasks were performed using the meta-analytic option in the StatsDirect® statistical program for those factors providing categorical data.

Selection Bias.

Many researchers have also questioned the validity of the published results due to the inconsistent selection of patients who undergo such a procedure (e.g. a patient eligible for a TOL may elect to have a ERCD or some institutions may have stricter criteria for a TOL). To address the potential of selection bias, this study performed a simple linear regression analysis in the SPSS (version 11.0)® statistical program, between the proportion of patients allowed a TOL and the proportion of patients with VD. As demonstrated by Rosen,[22] if selection bias is present, then the regression analysis would show a negative association between the proportion of patients allowed a TOL and proportion of patients with VD. For example, if a study is more selective as to whom they allow a TOL (lower proportion allowed a TOL), then one could expect a larger proportion of their population to have a VD.

Heterogeneity

In order to assess the heterogeneity between studies, the StatsDirect® statistical program was used to conduct the Breslow-Day test (i.e. Chi-Square test of Homogeneity), which determines how different a single OR (transformed by the natural log) is from the weighted average of all the study-specific odds ratios. If the statistic was found to have a level of significance (p) > 0.10 (which increases the probability of finding heterogeneity compared to a (p) of 0.05), then those studies being evaluated were considered to be homogeneous. However, if the studies for a particular factor were found to be heterogeneous (i.e. Breslow-Day test significance (p) < 0.10; and visually confirmed using forest plots), then an exploration of the finding was conducted using

subgroup analyses and the stratification of data, given that a sufficient number (i.e. three) of studies were available for such analyses.

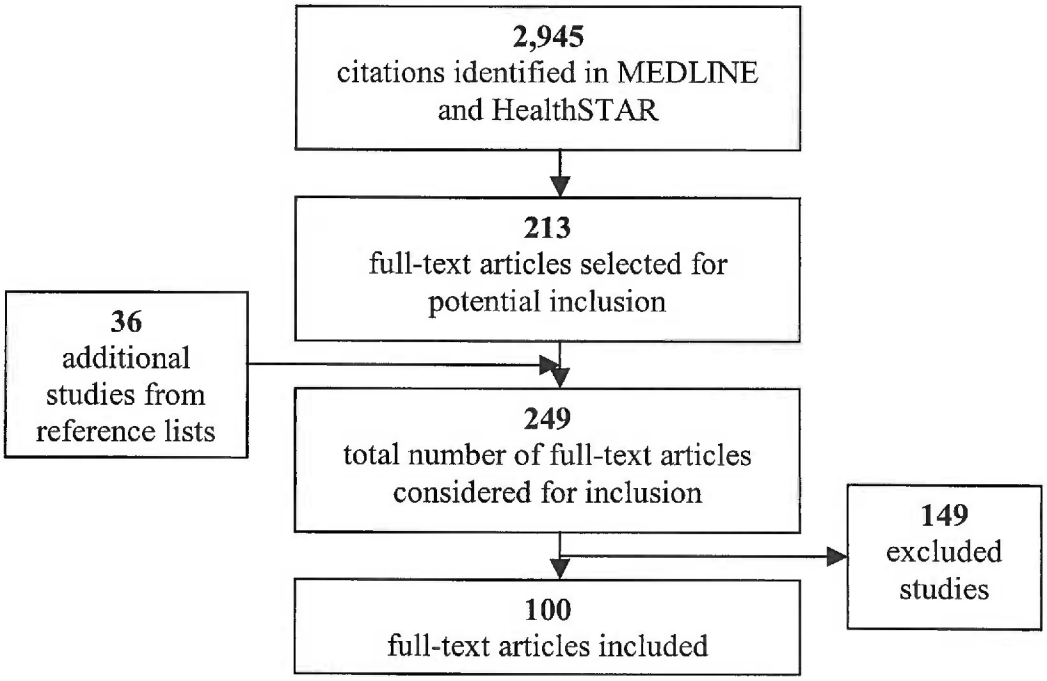
Theoretically, there are two types of subgroup analyses in a meta-analysis. One is the investigation of results defined by the study or patient characteristics and the other is the investigation of results while considering the subset of patients within the studies being pooled[45]. For the purposes of this study, it was decided that the focus would be placed on the investigation of heterogeneity based on the differences of study characteristics (e.g. study design, study quality, and sample size), due to inconsistent reporting of patient characteristics. Forest plots were used to assist with the visual identification of heterogeneity between studies.

Results:

1. *Literature Search and Selection of Articles*:

MEDLINE® and HealthSTAR® searches identified 2,945 titles and abstracts. Two hundred thirteen full-text articles were selected for inclusion. An additional 36 studies were identified in the reference lists of obtained articles, increasing the total number of full-text articles reviewed to 249. The full-text review of these articles and the re-application of inclusion criteria excluded 149 of the 249 (59.4%) (reasons for exclusion listed in Appendix D). The remaining 100 studies contained data pertinent to the investigation and were used to form the basis of the report (see Figure 1).

Figure 1. Study identification and selection flow chart



2. *Quality*:

Although the literature search resulted in a number of studies with information regarding certain predictive factors and their association to the outcome of a TOL following prior CD, the majority of these (82.0%) were given a rating of "poor" during the quality assessment. The most common indication for a "poor" rating was the lack of consideration and adjustment for confounding. Of the remaining studies, 13 were of "fair" quality and 5 were of "good" quality (see Appendix E, where the quality ratings for each of the 100 included studies are listed according to study design). Table 6 shows the overall number of studies grouped according to study design, receiving a "good," "fair," or "poor" quality rating.

Table 6. The number of good, fair, and poor quality studies according to study design

Study Design	Good Quality	Fair Quality	Poor Quality
RCT	1	1	0
Prospective Cohort	1	1	27
Retrospective Cohort	2	7	53
Case Control	0	2	2
Case series	1	2	0

Table 6 also shows that the cohort study was the most common type of study design used in the VBAC literature. The majority of these (i.e. 53 of 80) were retrospective, lending support to the idea that these studies are less expensive, quicker to complete, and overall easier to perform.

3. *Data Extraction*:

The 100 included full-text articles[9, 10, 17, 23-26, 28, 46-140] included a dramatic variation in the number of studies that examined each individual factor (range 0-61 studies). Table 7, lists those factors that were not analyzed in any fashion, due to the lack of any usable data. While two factors completely lacked any studies providing information, the remaining factors failed to meet the requirements listed in Table 5 (i.e. having at least three studies, of which they provide data in a fashion suitable for combination).

Table 7. Individual factors excluded from analysis, grouped by general categories

Category	Factors
Demographic	Race (1) SES (0)
Current Obstetric	Estimated Fetal Weight (3) Multiple Gestations (3) Breech/External Cephalic Version (3) Cervical dilation rate (2) Cervical effacement (2) Station (5) Bishop score (2) Labor duration (3) Maternal weight gain (3)
Non-Clinical	Insurance status (1) Hospital characteristics (2) Physician characteristics (0)

( ): number of studies for each factor

Table 8 provides a listing of the 18 factors for which the appropriate type and amount data was available for the calculation of summary estimates. Twelve of the 18 factors had categorical data suitable for combination, including three factors (e.g. previous VD order, number of prior CD, and prior CD indication) that provided more specific categorical data that allowed for more detailed analyses. In total, 18 categorical factor comparisons were made. Eight factors provided means and standard deviations suitable for combination, of which two (e.g. parity and birth weight) were also analyzed using categorical methods. In addition to providing the number of studies for each factor, Table 8 also provides the number of studies receiving "good" to "fair" quality ratings (see Appendix F for studies referenced by individual factor).

Table 8. Individual factors included in analysis, grouped by general categories

Category	Factors	
Demographic	Age (20/3)	
Past Obstetric	Gravidity (6/0) **Parity** (12/0) **Previous VD** (26/1) **Previous VD** order (10/4) **Before versus After** (7/1) **Before versus none before** (6/2) **After versus none after** (7/2) **Previous Cervical Dilation** (7/0)	**Number of prior CD** (22/1) **One versus More than One** (14/0) **One versus Two** (6/0) **One versus Three** (6/0) **Two versus Three** (8/0) **Prior CD Indications:** **Recurrent versus Non-recurrent** (61/2) **Recurrent versus Breech** (44/1) **Recurrent versus Fetal Distress** (41/1)
Current Obstetric	Gestational age (15/2) **Birth weight** (37/2) Cervical dilation (8/4) **Spontaneous labor** (26/0) **Induced labor** (26/0)	**Augmented labor** (21/2) **Oxytocin use (non-specified)** (25/0) **Epidural use** (16/1) Maternal height (5/0) Maternal weight (4/0)

Bold: are those factors for which *categorical* data was provided in a fashion suitable for combination
(/): number of studies for each factor / number of "good" and "fair" studies for each factor

Further review of the data revealed that 10 different patient populations (identified by year and location of study) were each used repeatedly by two to four different studies, so that in all 31 studies were found to have overlap. For the purposes of analysis, 19 studies were collapsed into seven general studies matched for their use of identical patient populations.

4. *Statistical Analysis*:

Summary Estimates

The random effects summary estimates (i.e. odds ratios for the likelihood of VD), calculated for those categorical factors providing data sufficient for combination are listed in Table 9. For example, 49 studies providing data on prior CD indication (recurrent versus non-recurrent) were pooled to provide an odds ratio for VBAC of 0.499

(95% CI 0.444, 0.562). The summary estimate indicates that those with a recurrent indication for their prior CD (e.g. cephalopelvic disproportion or failure to progress) were significantly less likely (i.e. half as likely) to have a VD compared to those with a non-recurrent indication for their prior CD (e.g. breech presentation, fetal distress, and all other indications).

Table 9. Classical random effects summary estimates for the individual factors providing categorical data

Factor	Number of Studies	Average VBAC proportion*	Odds Ratio (for VBAC)	95% CI
Past Obstetric				
Parity of one	3	55.9	**0.370**	**0.193, 0.710**
Number of prior CD				
One prior CD (versus more than one)	14	80.2	**1.532**	**1.358, 1.729**
One prior CD (versus two prior CD)	6	82.8	**1.526**	**1.212, 1.921**
One prior CD (versus three prior CD)	6	82.8	1.234	0.913, 1.668
Two prior CD (versus three prior CD)	8	75.1	1.160	0.726, 1.854
Previous VD	19	88.8	**3.182**	**2.583, 3.921**
Previous VD order				
Before prior CD (versus after)	7	80.5	**0.382**	**0.303, 0.483**
Before prior CD (versus none before)	6	82.6	**1.473**	**1.247, 1.741**
After prior CD (versus none after)	7	93.5	**4.330**	**3.108, 6.031**
Indication of prior CD				
Recurrent (versus non-recurrent)	49	64.5	**0.499**	**0.444, 0.562**
Breech (versus recurrent)	34	84.1	**2.888**	**2.379, 3.507**
Fetal Distress (versus recurrent)	32	72.8	**1.513**	**1.340, 1.648**
Previous Cervical Dilation <4cm	3	68.9	0.823	0.456, 1.484
Current Obstetric				
Spontaneous labor	19	76.2	**1.884**	**1.421, 2.497**
Oxytocin use (non-specified)	18	67.6	**0.444**	**0.315, 0.625**
Augmented labor	16	79.5	**0.574**	**0.346, 0.955**
Epidural use	12	75.8	0.612	0.330, 1.138
Birth weight <4000g	12	73.3	**2.274**	**1.713-3.019**

Bold = significant difference; * average VBAC proportion in those with that factor

Those factors found to be significantly associated with an increased likelihood of a successful TOL (i.e. VD) include having only one prior CD (versus more than one prior CD), having a previous VD (including both those having a VD before and those having a VD after prior CD), a prior CD indication of breech or fetal distress, having a spontaneous onset of labor, and having a infant weighing less than 4000g. Using the inverse odds ratio, it can be concluded that those with parity more than one, a previous VD after prior CD compared to before prior CD, a non-recurrent versus a recurrent prior CD indication, not using any oxytocin, and not being augmented were also significantly more likely to be successful in their TOL.

These summary estimates were all similar to those estimates calculated using the WINBUGS® statistical program (see Appendix G), with the exception of augmented labor factor. The discrepancy between the two calculations (StatsDirect® - OR 0.574, WINBUGS® - OR 1.423) was found to be due to a single large study,[114] whose odds ratio was closer to the WINBUGS® estimate. Because the study was so large, the within-study variance was relatively small compared to the large amount of between-study variance for the factor. The disproportion in variance resulted in the minimal importance of the study's results in the random effects process employed by the StatsDirect program®. The finding was not reproduced by WINBUGS® because of the fact that the program uses hypothetical simulations of the data for the creation of their summary estimates.

Available estimates provided by good and fair quality studies for the categorical factors are listed in Table 10. In general, the findings from the random effects modeling

were similar (i.e. same direction and trend of odds ratios) to the individual findings of these quality studies, providing support for the validity of the summary estimates.

Table 10. Effect estimates provided by good and fair quality studies, for the individual factors for which categorical summary estimates were calculated

Factor	Author (year)	Adjusted OR for VBAC	95% CI, p-value
Past Obstetric			
Number of prior CD*	Pickhardt (1992)	**0.43**	**p<0.05**
Previous VD	McNally (1999)	**27.78**	**3.85-200**
Previous VD order			
Before prior CD (versus after)	Caughey (1998)	**0.287**	**0.164-0.526**
Before prior CD (versus none before)	Flamm (1997)	**1.53**	**1.12-2.10**
	Weinstein (1996)	**1.8**	**1.1-3.1**
After prior CD (versus none after)	Flamm (1997)	**3.39**	**2.25-5.11**
	Macones (2001)	**7.69**	**3.23-20**
Prior CD Indication			
Recurrent versus Non-recurrent	Flamm (1997)	**0.518**	**0.426-0.633**
	Weinstein (1996)	0.8	0.3-2.0
Breech versus Recurrent	Weinstein (1996)	1.9	1.0-3.6
Fetal Distress versus Recurrent	Weinstein (1996)	1.05	0.4-2.6
Current Obstetric			
Augmented Labor	Macones (2001)	**0.47**	**0.25-0.88**
	Stronge (1996)	NR	NS
Epidural use	McNally (1999)	0.26	0.06-1.12
Birth weight <4000g	Weinstein (1996)	1.053	0.2-5.882
	Zelop (2001)	**1.695**	**1.299-2.222**

Bold: significant; NR = not reported; NS = not significant
*Regression analysis of increasing number of prior CD and proportion with VD

The calculation of summary mean estimates for those with a successful TOL (i.e. vaginal birth) and those with a failed TOL are listed in Table 11. The differences in means that were found to be significant are for the factors of maternal age, birth weight, and maternal weight. Those with a VD were nearly a year younger and weighed 1.57 kg less, and also had an infant that weighed 177.3 grams less compared to those with a failed

TOL. The estimates of the "good" and "fair" quality studies (see Table 12) support the differences and trends observed in the summary mean comparisons.

Table 11. Bayesian random effects summary mean estimates for the individual factors providing continuous data

Factor	Number of Studies	VBAC - Mean	Failed TOL – Mean	Difference between Means (VBAC–FTOL)
Demographic				
Age (yrs)	13	29.11	30.03	**-0.92 (-1.748, -0.211)**
Past Obstetric				
Gravidity	5	2.94	2.483	0.4564 (-0.744, 1.718)
Parity	5	2.275	1.735	0.5395 (-0.794, 1.683)
Current Obstetric				
Gestational age (wks)	10	38.79	39.47	-0.6838 (-1.627, 0.125)
Admission cervical dilation (cm)	3	3.647	2.23	1.417 (-0.304, 2.894)
Birth weight (g)	14	3288	3465	**-177.3 (-178.1, -176.6)**
Maternal height (cm)	5	157.1	156.0	1.156 (-0.178, 2.299)
Maternal weight (kg)	3	67.8	69.37	**-1.573 (-3.294, -0.096)**

Bold = significant difference

Table 12. Effect estimates provided by good and fair quality studies, for the individual factors for which mean summary estimates were calculated

Factor	Author (year)	Adjusted* OR for VBAC	95% CI, p-value
Demographic			
Maternal Age	Flamm (1997)	**2.58** (<40 yrs)	1.55, 4.3
	McNally (1999)	1.18 (per yr of age)	0.98, 1.40
	Weinstein (1996)	0.9 (>37yr)	0.5, 1.7
Current Obstetric			
Gestational Age	Pickhardt (1992)	**0.81**	p<0.05
	Zelop (2001)	**0.67** (>40wks GA, spontaneous)	0.56, 0.83
	Zelop (2001)	**0.67** (>40wks GA, induced)	0.45, 0.91
Cervical Dilation	Flamm (1997)	**2.16** (>4cm)	1.66, 2.82
	Macones (2001)	**1.87**	1.14, 3.23
	Pickhardt (1992)	**1.62**	p<0.05
	Stronge (1996)	NR	NS
Birth weight <4000g	Weinstein (1996)	1.053	0.2, 5.882
	Zelop (2001)	**1.695**	**1.299, 2.222**

Bold: significant; *OR adjusted for various confounding factors, which varied from study to study

The visual examination of the categorical factors involved the use of forest plots that displayed the distribution of estimates and confidence intervals for each study, in order of sample size (i.e. smallest study to largest study). Figure 2 illustrates the forest plot for the factor of prior CD indication (recurrent versus non-recurrent). The clear diamond represents the random effect summary estimate and the accompanying horizontal line represent the confidence interval (0.499 and 95% CI 0.444, 0.562, respectively) for the factor. Visually, the figure also describes the relative sample size of each study by the size of the respective black square (and relative position on the vertical axis), as well as the range of each study's confidence interval by the width of the accompanying horizontal line. The forest plot visually demonstrates no obvious relationship or pattern between the odds of vaginal birth after CD for those with a recurrent versus non-recurrent prior CD indication and the size of the study's sample population.

In Figure 3, for previous VD (before versus none), there appears to be a trend for the estimates by sample size. Although the visual interpretation is subjective, it does show that as the sample size increases so do the odds of a VD for those with a previous VD before prior CD compared to those without one (see Appendix H – forest plots for remaining categorical factors).

Figure 2. Forest plot according to sample size (i.e. smallest to largest), for the factor of prior CD indication: recurrent versus non-recurrent

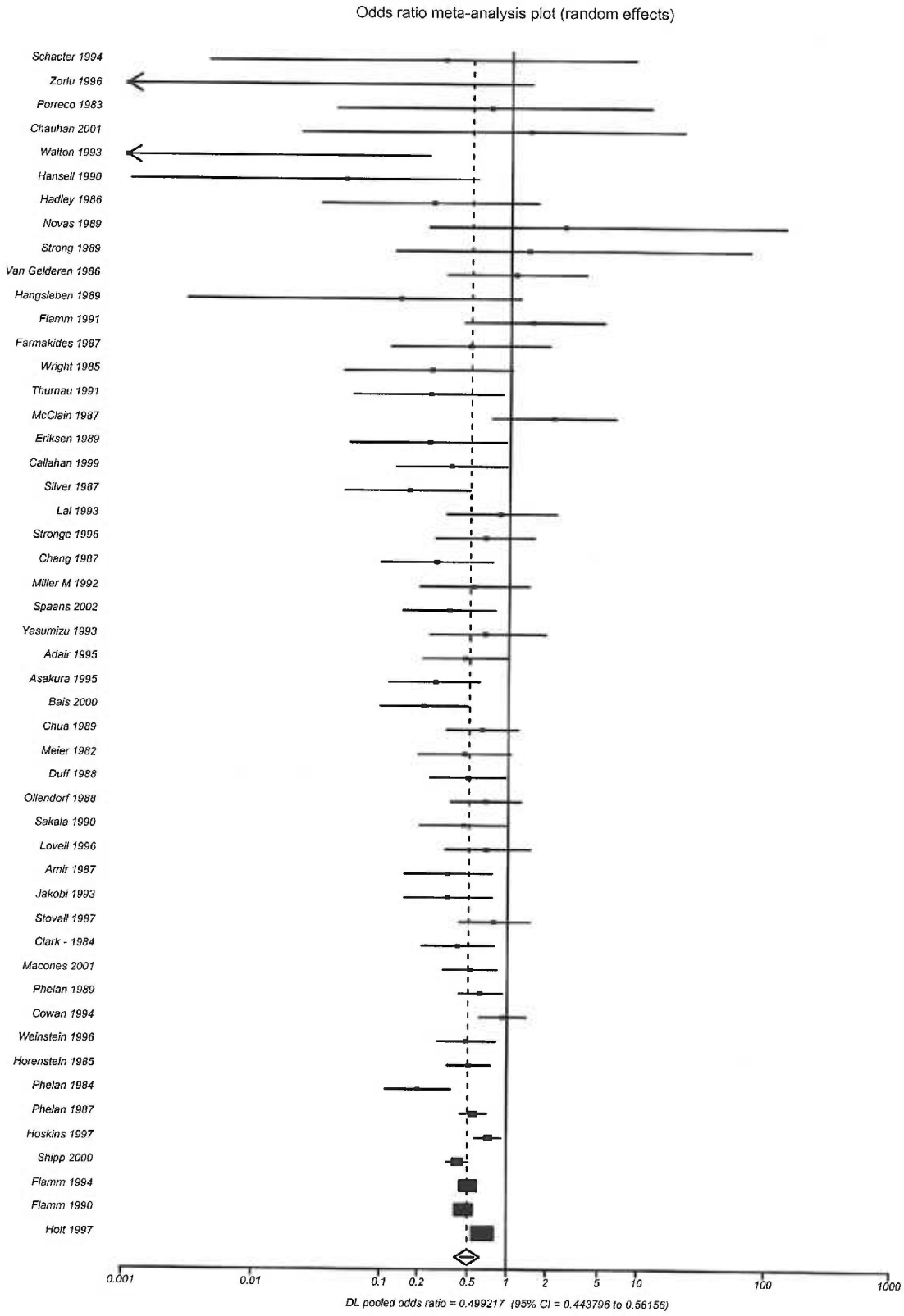
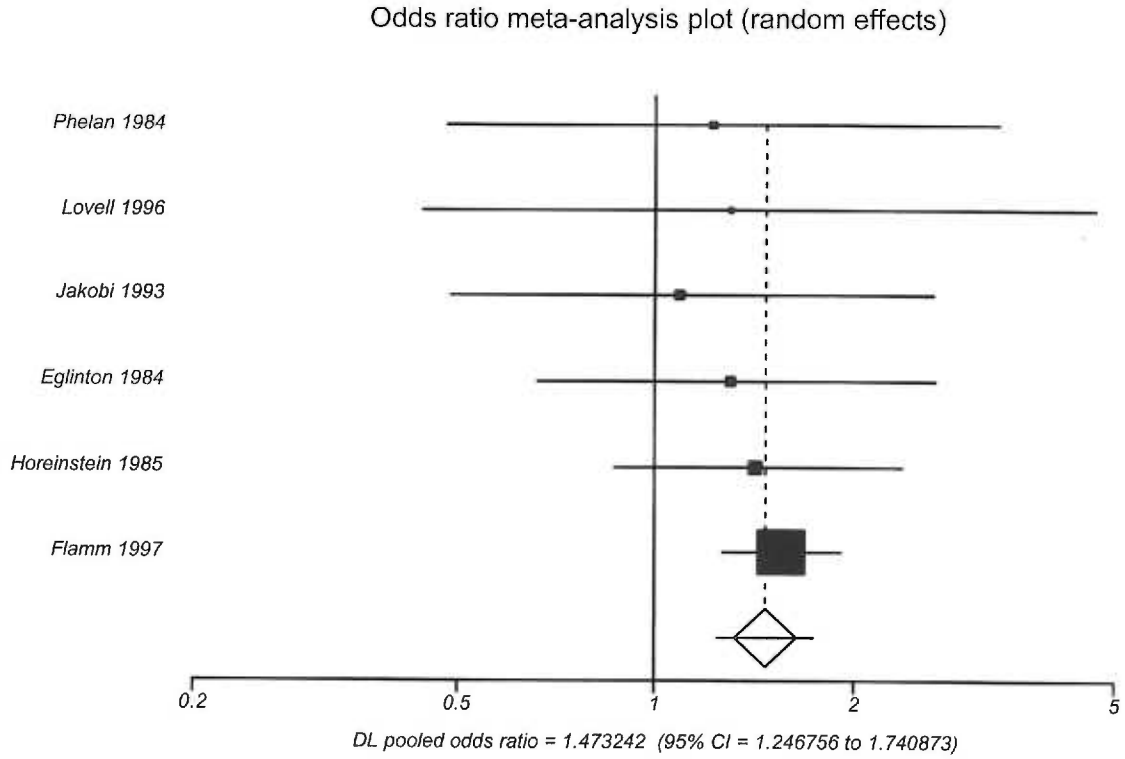


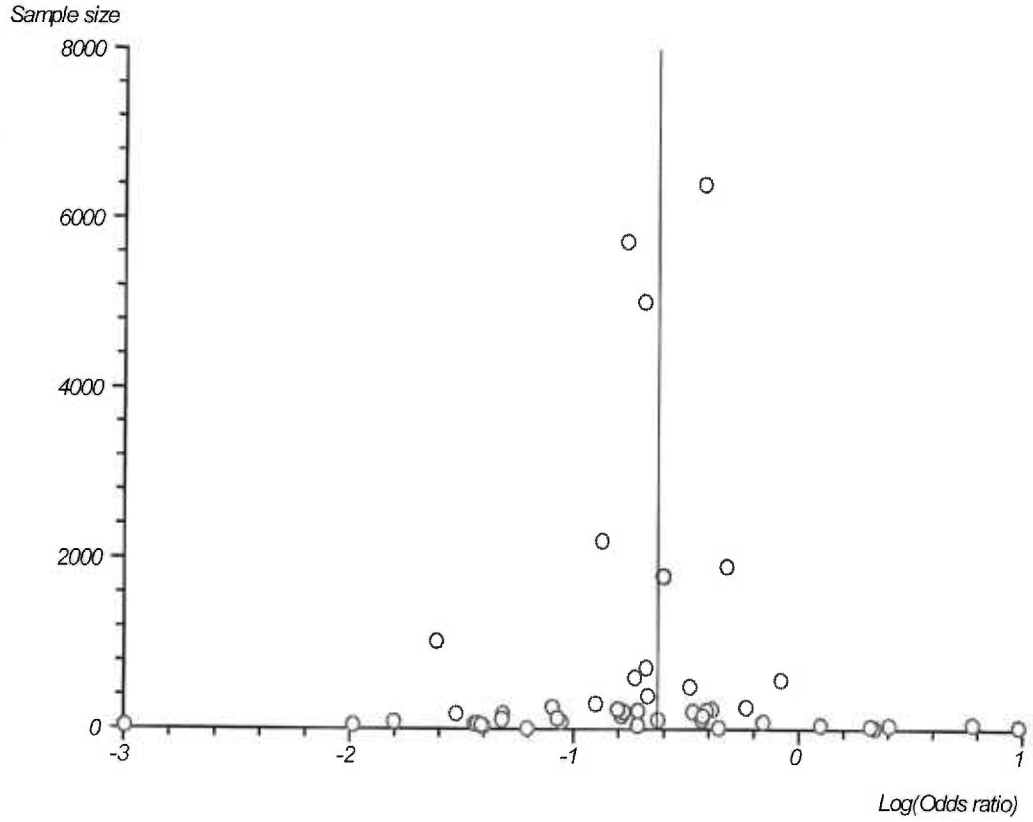
Figure 3. Forest plot according to sample size (i.e. smallest to largest), for the factor of previous VD: before prior CD versus none



Publication Bias

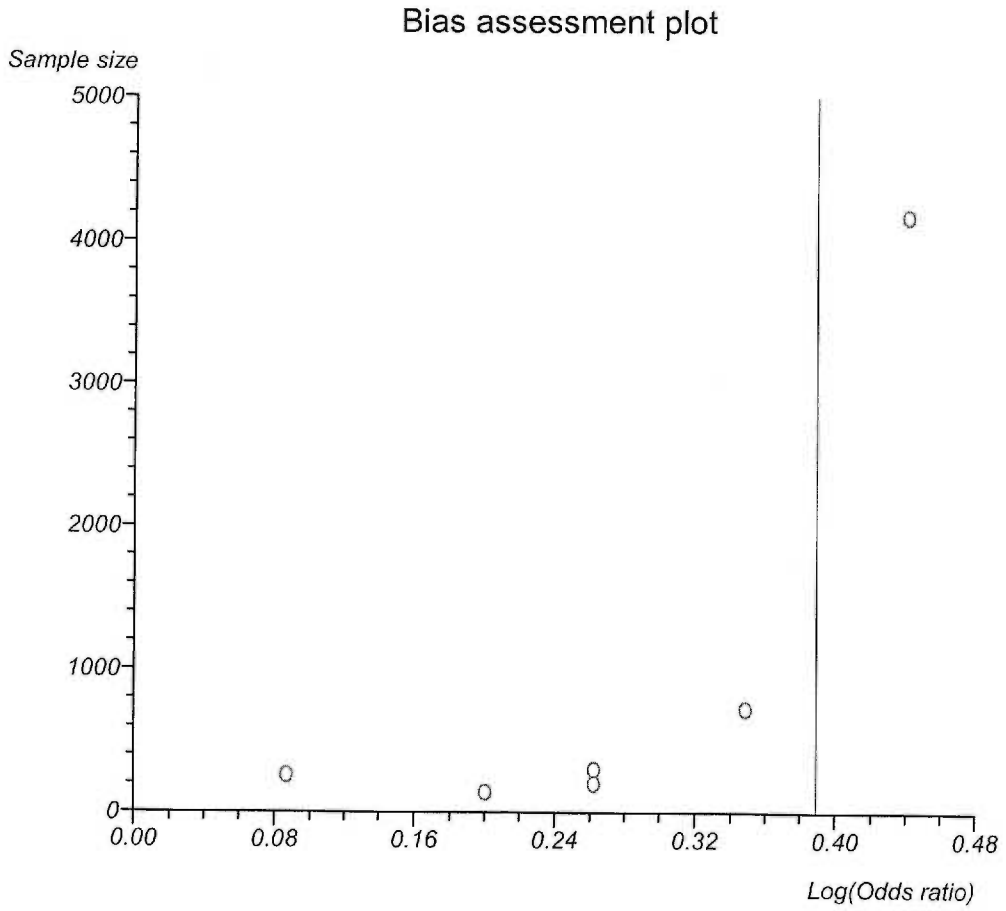
Multiple funnel plots were required in order to identify publication bias because of the fact that the investigation looked at multiple factors and their association to TOL outcome. As an example, Figure 4 demonstrates the funnel plot for the prior CD indication (recurrent versus non-recurrent) factor. Visually, the figure shows no obvious asymmetry of the distribution of studies (plotted using the respective study size and log odds ratios), indicating the lack of publication bias for the factor. The regression analysis of the normalized effect versus precision statistically supported the lack of publication bias, by showing that the intercept did not significantly differ from zero (-0.334, 95% CI = - 0.877, 0.209, p = 0.2219).

Figure 4. The assessment of publication bias using a funnel plot, for the factor of prior CD indication: recurrent versus non-recurrent



Another example of the assessment of publication bias using a funnel plot is seen in Figure 5, for the factor of previous VD (before prior CD versus none). Unlike the previous forest plot for prior CD indication (recurrent versus non-recurrent), the forest plot does appear to be asymmetrical, with smaller studies tending to show smaller effect estimates. The subjective interpretation was supported statistically by the regression analysis of the normalized effect versus precision, which found that the intercept did significantly differ from zero (-0.660, 95% CI = -1.053, -0.267, p = 0.010).

Figure 5. The assessment of publication bias using a funnel plot, for the factor of previous VD: before prior CD versus none



Since the visualization of funnel plots (see Appendix I – funnel plots for remaining categorical factors) requires that subjective judgments be made, the conclusion on the presence of publication bias was based heavily on the statistical regression analyses between study size and study estimates (see Table 13). Publication bias was detected for the factors of one prior CD (versus more than one), previous VD before prior CD (versus none before), and the augmentation of labor.

Table 13. The assessment of publication bias using regression analysis

Factor	Intercept	95% Confidence Interval	p-value
Past Obstetric			
Parity of one	NA	NA	NA
Number of prior CD			
One prior CD (versus more than one)	**-0.763**	**-1.488, -0.038**	**0.041**
One prior CD (versus two prior CD)	-0.878	-2.387, 0.631	0.182
One prior CD (versus three prior CD)	-0.920	-3.395, 1.555	0.133
Two prior CD (versus three prior CD)	0.234	-2.523, 2.991	0.825
Previous VD	0.629	-0.158, 1.417	0.110
Previous VD order			
Before prior CD (versus after)	-0.185	-1.981, 1.612	0.790
Before prior CD (versus none before)	**-0.660**	**-1.053, -0.267**	**0.010**
After prior CD (versus none after)	0.173	-1.563, 1.909	0.796
Indication of prior CD			
Recurrent (versus non-recurrent)	0.334	-0.87, 0.209	0.222
Breech (versus recurrent)	-0.035	-0.821, 0.750	0.927
Fetal Distress (versus recurrent)	-0.062	-0.525, 0.400	0.784
Previous Cervical Dilation <4cm	NA	NA	NA
Current Obstetric			
Spontaneous labor	0.500	-1.076, 2.076	0.511
Oxytocin use (non-specified)	-0.346	-2.488, 1.795	0.734
Augmented labor	**-2.709**	**-4.316, -1.102**	**0.0028**
Epidural use	-2.548	-7.065, 1.896	0.228
Birth weight <4000g	1.504	-0.325, 3.337	0.095

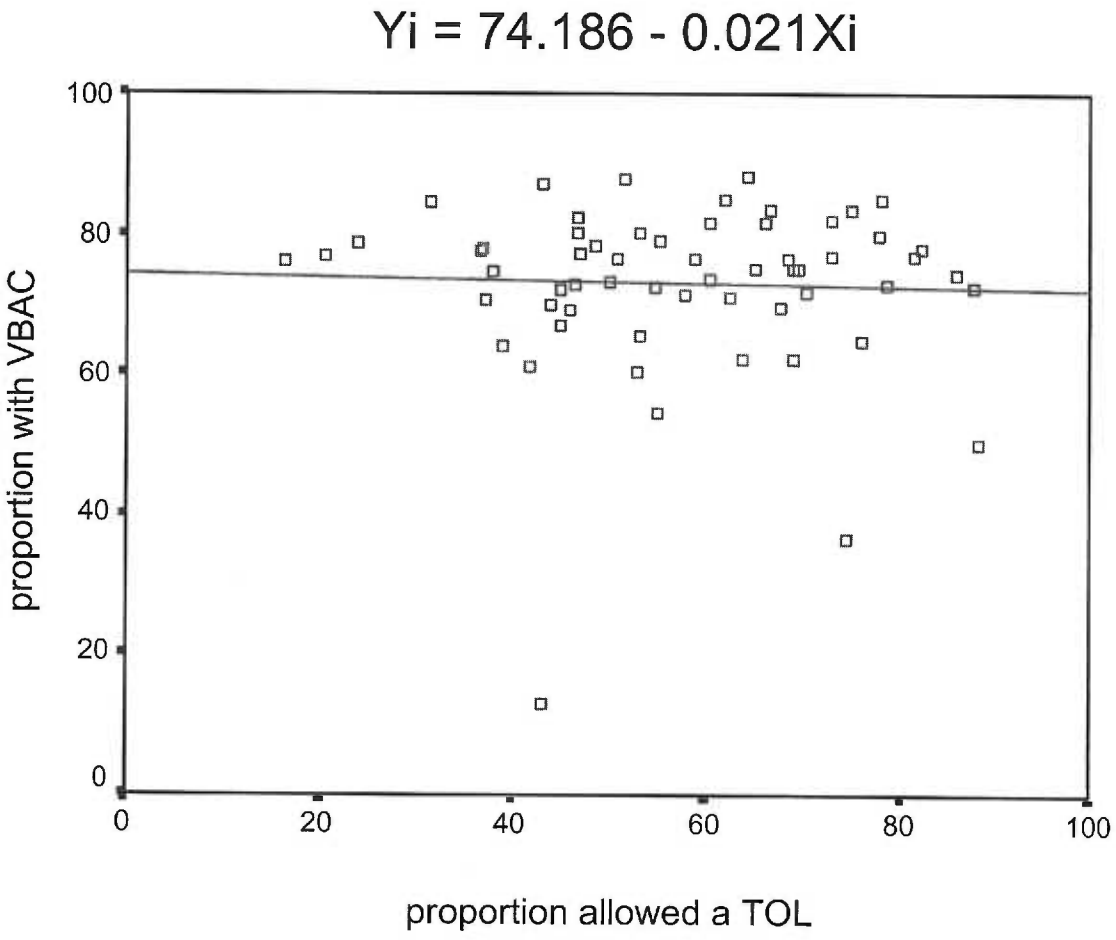
NA: not available, due to an insufficient number of studies for analysis
Bold: publication bias present according to regression analysis

Selection Bias

Rosen's[22] technique for investigating the presence of a selection or work-up bias was performed using the data from 63.0% or 63 of the 100 studies providing the data necessary for the calculation. The analysis of these studies using SPSS (version 11.0)® resulted in a regression equation ($Y_i = 74.186 - 0.021X_i$, where Y_i represents the proportion of those with VBAC for a given proportion allowed a TOL (X_i)) with an F-test and p-value of 0.051 and 0.823 respectively, demonstrating that there was no significant association between the proportion of those allowed a TOL and the proportion of VD in

those attempting a TOL. The lack of an association suggests that a selection or work-up bias was not likely in those studies being evaluated. Figure 6 supported the finding by illustrating the non-uniform distribution of study points and the nearly flat regression equation (i.e. fitted) line.

Figure 6. The assessment of selection bias using simple linear regression



Heterogeneity

The identification of heterogeneity between studies (i.e. the Breslow-Day test with a level of significance (p) <0.10) using the StatsDirect® statistical program revealed

that 61.1% (i.e. 11 out of 18) of the categorical factors were found to have studies that were heterogeneous (see Table 14).

Table 14. The assessment of study heterogeneity using the Breslow-Day test statistic

Factor	Breslow-Day test statistic	df	p-value
Past Obstetric			
Parity of one	**10.750**	**2**	**0.005**
Number of prior CD			
One prior CD (versus more than one)	15.442	13	0.281
One prior CD (versus two prior CD)	7.472	5	0.188
One prior CD (versus three prior CD)	6.841	5	0.233
Two prior CD (versus three prior CD)	**14.766**	**7**	**0.039**
Previous VD	**28.384**	**18**	**0.056**
Previous VD order			
Before prior CD (versus after)	7.463	6	0.280
Before prior CD (versus none before)	1.272	5	0.938
After prior CD (versus none after)	8.397	6	0.210
Indication of prior CD			
Recurrent (versus non-recurrent)	**127.520**	**48**	**<0.001**
Breech (versus recurrent)	**88.839**	**33**	**<0.001**
Fetal Distress (versus recurrent)	**37.149**	**31**	**<0.001**
Previous Cervical Dilation <4cm	3.427	2	0.180
Current Obstetric			
Spontaneous labor	**123.654**	**18**	**<0.001**
Oxytocin use (non-specified)	**132.453**	**17**	**<0.001**
Augmented labor	**218.252**	**15**	**<0.001**
Epidural use	**129.340**	**11**	**<0.001**
Birth weight <4000g	**66.439**	**11**	**<0.001**

Bold: significant heterogeneity present

The exploration of study heterogeneity using subgroup analyses based on study design, study quality, and study sample size was able to demonstrate that these characteristics did explain in part, some of the observed heterogeneity (e.g. subgroup analysis where the Breslow-Day statistic became non-significant, the summary estimate became significant or non-significant, or the summary estimate was found to be in the

opposite direction – bold face type in Tables 16, 17, and 18). Although the partial contribution to study heterogeneity appears evident, Table 15 shows that none of the study characteristics were able to fully explain the observed heterogeneity.

Table 15. The findings of the heterogeneity exploration according to study design, study quality, and study sample size

Factor	Study Design	Study Quality	Study Sample Size
Past Obstetric			
Parity of one	NA	NA	NA
Number of prior CD (two versus three)	N	N	N
Previous VD	Y/N	Y/N	N
Prior CD Indication (recurrent versus non-recurrent)	N	Y/N	N
Prior CD Indication (breech versus recurrent)	N	Y/N	N
Prior CD Indication (fetal distress versus recurrent)	Y/N	Y/N	N
Current Obstetric			
Spontaneous labor	Y/N	N	N
Oxytocin use (non-specified)	Y/N	N	N
Augmented labor	Y/N	N	Y/N
Epidural use	Y/N	N	N
Birth weight <4000g	N	N	Y/N

NA: not applicable due to inappropriate number of studies per subgroup
N: no explanation for observed heterogeneity
Y/N: some explanation for observed heterogeneity
Y: explanation for observed heterogeneity

In order to illustrate how the conclusions in Table 15 were reached, the remainder of this section details the subgroup analyses for the factor of prior CD indication (recurrent versus non-recurrent). With regard to study design (see Table 16), there were four subgroups considered for analysis. Both the case control and case series design subgroups had two studies each, making them ineligible for pooling. The remainder of the 45 studies was comprised of 16 prospective design studies and 29 retrospective design studies. The summary estimates for the prospective design and retrospective

design studies were 0.542 (95% CI 0.462, 0.635) and 0.441 (95% CI 0.365, 0.534), respectively. The comparison of these design specific summary estimates to each other and to the overall summary estimate of 0.499 (95% CI 0.444, 0.562), revealed no significant differences by study design. Additionally, since the heterogeneity remained significant for both study design subgroups (i.e. prospective Breslow-Day = 33.421 (df =15), p = 0.004 and retrospective Breslow-Day = 87.876 (df = 28), p <0.001), stratifying by the definition of study design used does not appear to explain the majority of the heterogeneity between studies.

Table 16. The exploration of study heterogeneity, using subgroup analyses according to study design

Factor	Study Design		Overall Summary Estimate
	Prospective	Retrospective	
Past Obstetric			
Parity of one	NA [0]	0.370 (0.193,0.710) [3] - BD p=0.005	0.370 (0.193,0.710)
Number of prior CD (two vs three)	1.233 (0.725,2.096) [6] - BD p=0.016	NA [2]	1.160 (0.726,1.854)
Previous VD	2.551 (2.202,2.955) [6] - **BD p=0.887**	3.700 (2.632,5.199) [11] - BD p=0.049	3.182 (2.583,3.921)
Prior CD Indication (recurrent vs non-recurr)	0.542 (0.462,0.635) [16] - BD p=0.004	0.441 (0.365,0.534) [29] - BD p<0.001	0.499 (0.444,0.562)
Prior CD Indication (breech vs recurr)	2.840 (2.037,3.958) [10] - BD p<0.001	2.979 (2.340,3.794) [21] - BD p=0.009	2.888 (2.379,3.507)
Prior CD Indication (fetal distress vs recurr)	1.527 (1.358,1.717) [10] - **BD p=0.760**	1.552 (1.308,1.842) [19] - BD p=0.069	1.513 (1.340,1.648)
Current Obstetric			
Spontaneous labor	1.811 (1.268,2.586) [5] - **BD p=0.101**	1.722 (1.202,2.614) [12] - BD p<0.001	1.884 (1.421,2.497)
Oxytocin use (non-specified)	0.352 (0.226,0.550) [8] - BD p<0.001	0.563 (0.292,**1.085**) [10] - BD p<0.001	0.444 (0.315,0.625)
Augmented labor	0.396 (0.202,0.777) [4] - BD p<0.001	0.806 (0.470,**1.382**) [10] - BD p<0.001	0.574 (0.346,0.955)
Epidural use	**1.219** (0.650,2.283) [4] - BD p<0.001	0.592 (0.202,1.735) [6] - BD p<0.001	0.612 (0.330,1.138)
Birth weight <4000g	NA [2]	2.013 (1.508,2.688) [8] - BD p<0.001	2.274 (1.713,3.019)

NA: not applicable due to inappropriate number of studies per subgroup
Cell: summary estimate odds ratio (95% confidence interval), [number of studies] – Breslow-Day p-value

The examination of subgroups based on quality ratings (see Table 17) for the factor of prior CD indication (recurrent versus non-recurrent), led to the comparison of six studies rated as being "good" or "fair" quality and 43 studies rated as "poor". The summary estimates for the "good" and "fair" quality studies and the "poor" quality studies were 0.520 (95% CI 0.386, 0.701) and 0.498 (95% CI 0.440, 0.565), respectively. Once again the comparison of each subgroup with each other and with the overall summary estimate of 0.499 (95% CI 0.444, 0.562), indicated no significant differences by study quality. While the "poor" quality studies remained heterogeneous

Table 17. The exploration of study heterogeneity, using subgroup analyses according to study quality ratings

Factor	Study Design		Overall Summary Estimate
	"Good/Fair"	"Poor"	
Past Obstetric			
Parity of one	NA [1]	NA [2]	0.370 (0.193,0.710)
Number of prior CD (two vs three)	NA [0]	1.160 (0.726,1.854) [8] - BD=0.039	1.160 (0.726,1.854)
Previous VD	NA [2]	2.996 (2.448,3.667) [17] - **BD p=0.437**	3.182 (2.583,3.921)
Prior CD Indication (recurrent vs non-recurr)	0.520 (0.386,0.701) [6] - **BD p=0.350**	0.499 (0.444,0.562) [43] - BD p<0.001	0.499 (0.444,0.562)
Prior CD Indication (breech vs recurr)	4.972 (2.817,8.776) [3] - **BD p=0.677**	2.769 (2.264,3.387) [29] - BD p<0.001	2.888 (2.379,3.507)
Prior CD Indication (fetal distress vs recurr)	NA [2]	1.513 (1.379,1.661) [30] - **BD p=0.169**	1.513 (1.340,1.648)
Current Obstetric			
Spontaneous labor	NA [2]	1.989 (1.458,2.713) [17] - BD p<0.001	1.884 (1.421,2.497)
Oxytocin use (non-specified)	NA [0]	0.444 (0.315,0.625) [18] - BD p<0.001	0.444 (0.315,0.625)
Augmented labor	NA [1]	0.609 (0.364,1.020) [15] - BD p<0.001	0.574 (0.346,0.955)
Epidural use	NA [1]	0.667 (0.351,1.267) [11] - BD p<0.001	0.612 (0.330,1.138)
Birth weight <4000g	NA [2]	2.288 (1.721,3.043) [10] - BD p<0.001	2.274 (1.713,3.019)

NA: not applicable due to inappropriate number of studies per subgroup
Cell: summary estimate odds ratio (95% confidence interval), [number of studies] – Breslow-Day p-value

(Breslow-Day = 122.366 (df = 42), p<0.001), the Breslow-Day test for the "good" and "fair" quality studies indicated that they were homogeneous (Breslow-Day = 5.573 (df = 5), p = 0.35). From the finding, one could speculate that only a minor portion of the study heterogeneity may be explained by study quality.

The lack of a clear trend between study sample size and estimates for the factor of prior CD indication (recurrent versus non-recurrent) visually indicates that sample size does not contribute to study heterogeneity (see Figure 2). The regression analysis for publication bias (see Table 13), which regressed study sample size on study estimates, confirmed the lack of an association with a p-value of 0.222. Altogether, these finding indicate that while study design and sample size cannot explain the observed heterogeneity for the factor of prior CD indication (recurrent versus non-recurrent), study quality may be a possible explanation for some of the variability (see Table 18).

Table 18. The assessment of the contribution of sample size in study heterogeneity

Factor	Forest Plot Trend*	Regression p-value
Past Obstetric		
Parity of one	NA	NA
Number of prior CD (two versus three)	no apparent trend	0.825
Previous VD	no apparent trend	0.110
Prior CD Indication (recurrent versus non-recurrent)	no apparent trend	0.222
Prior CD Indication (breech versus recurrent)	no apparent trend	0.927
Prior CD Indication (fetal distress versus recurrent)	no apparent trend	0.784
Current Obstetric		
Spontaneous labor	no apparent trend	0.511
Oxytocin use (non-specified)	no apparent trend	0.734
Augmented labor	no apparent trend	**0.0028**
Epidural use	no apparent trend	0.228
Birth weight <4000g	possible trend	0.095

NA: not applicable due to inappropriate number of studies; * subjective interpretation of visual trend

Discussion/Limitations:

Among many reasons, rising CD rates and the underutilization of VBAC attempts made it the objective of this study to address the question of *"what factors predict the route of delivery for patients with a prior CD who undergo a trial of labor?"*

The first step of *completing a systematic literature review in order to clearly identify those factors that have been associated with either a successful or failed TOL*, identified 100 studies for inclusion. These studies varied in terms of quality, with 82% (i.e. 82 out of 100 studies) being rated as "poor". The most common indication for this rating was the lack of accounting for potential confounders, which questions the validity of these studies. The finding that only 37.5% (12 out of 32) of the proposed factors had sufficient data to calculate either summary estimates or summary mean estimates, clearly demonstrates the limitations of the past 22 years of VBAC literature and supports the need for more research regarding these potentially influential factors.

The second step of *using random effects modeling in order to calculate summary estimates for each factor*, revealed several findings. Unlike previous reviews and meta-analyses, this study was able to produce methodologically valid results by accounting for study heterogeneity through random effects modeling. Consistent with previous studies, it was found that numerous categorical factors were significantly associated with VD, including: parity greater than one, only one prior CD, previous VD (especially VD after prior CD), non-recurrent indication for prior CD (including breech presentation and fetal distress), spontaneous labor, no oxytocin use (including for augmentation), and a birthweight less than 4000g. The only unexpected finding from these analyses was that while having one prior CD (versus two prior CD) was significantly associated with VD

(1.526, 95% CI 1.212, 1.921), having one prior CD (versus three prior CD) was not. After reviewing the, it was reasoned that the result was due to the lack of power to find a difference, as supported by the fact that the number of subjects with three prior CD was fewer than those with two prior CD, and the finding that having one prior CD (versus three prior CD) trended toward an increased likelihood of VD (1.234, 95% CI 0.913, 1.668).

In comparing the results from this study to the meta-analysis conducted by Rosen,[22] it was found that the overall summary estimates were similar (see Table 19). While reassuring, the finding may cause one to question the utility of using the more time consuming and statistically complex random effects model, if the fixed effects model produces the same estimates. In the end however, the methodologic rigor of accounting for study heterogeneity (which is present according to the heterogeneity analyses of this study) through the use of a random effects model provides more confidence and support for the validity of the estimates provided by this study.

Table 19. A comparison of fixed and random effects summary estimates (95% CI)

Factor	Rosen Meta-Analysis (fixed effects)	Current Meta-Analysis (random effects)
Recurrent Indication	0.5 (0.5, 0.6)	0.488 (0.444-0.562)
Breech Indication	2.1 (1.8, 2.3)	2.888 (2.379-3.507)
Previous VD	2.1 (1.7, 2.5)	3.182 (2.583-3.921)
More than one prior CD	0.7 (0.5, 0.9)	0.652 (0.578-0.736)
Oxytocin use	0.3 (0.3, 0.4)	0.440 (0.315-0.625)

For those factors that had studies providing continuous data, investigators felt that in addition to providing a comparison between those who succeeded and those who failed at a TOL for a specific factor, summary mean calculations would also for the estimation

of potential categorical cut-points for future research using these factors. Of the eight factors with continuous data, only three factor comparisons (e.g. age, birthweight, and maternal weight) were found to have significantly different values between those who succeeded and those who failed in their TOL. For example, those with VBAC had a mean infant birthweight of 3288g, while those who failed their TOL had a mean infant birthweight of 3465g. Although statistically significant, the difference of 177.3g (95% CI 176.6, 178.1) in birthweight did not add much in terms of clinical relevance, especially since it is already accepted clinically that women with larger infants tend to fail in their TOL more often than those with smaller infants. The small difference in birthweight of only 177.3g also made the objective of establishing a new categorical cut-point difficult. The presence of this finding with all three factors supports the need for careful interpretation of any results, while keeping in mind that statistical significance does not necessarily entail clinical significance.

With regard to meta-analyses, many have stated that the quality of the final product is largely determined by the quality of the individual studies that are included. While it would have been ideal to calculate the summary estimates from only "good" or "fair" quality studies, the restriction would have dramatically reduced the number of factors that provided a sufficient number of studies for statistical combination. Although our final summary estimates did include "poor" quality studies, the validity of these estimates is supported by their similarity to the those estimates of the individual "good" and "fair" quality studies, as well as the similarity to the those estimates calculated through subgroup analyses based on quality. While this similarity lends confidence, one could alternatively interpret these findings to indicate the inadequacy of the quality rating

system utilized in this study. Although adapted from the previous validated USPSTF and York criteria, this study's rating system has yet to be tested. While this finding could also represent the misapplication of this study's rating system, it is less likely considering the discussion and review of the applied criteria by two investigators. Nonetheless, these possibilities support the need for caution in the interpretation and application of the findings in this study.

The wide range in the numbers of studies used for the calculation of summary estimates (e.g. three to 61) raises the question of the level of confidence that can be applied to each estimate. For example, one may feel more confident accepting a summary estimate based on 61 studies compared to one based only on three studies. In the end, the differences in the number of studies used for calculating each estimate needs to be carefully considered.

As a secondary gain of this random effects modeling process, it was established that both the Classical and Bayesian techniques had their advantages and disadvantages. While the estimates of these methods tend to be relatively similar (especially with a large number of studies), several aspects of each method make them truly distinct from each other. Perhaps the most striking difference lies in the fact that the Bayesian method relies heavily on the existing knowledge base. Although the use of these "informative priors" can be seen as an advantage, in that individual studies borrow strength from one another, it likewise can be seen as a disadvantage in that rather than being objective, the estimates are often considered to be "subjective" probabilities. In addition to this, other limitations of the Bayesian method include the difficulties of eliciting the prior beliefs and the lack of established guidelines to help with such computationally complex and time-consuming

procedures[40]. Even though this study could not find strong evidence to support the use of one method over another, the final decision to use the Classical approach mainly for the objective nature of its estimates and because it allowed for a more comprehensive meta-analytic approach, including the exploration of study heterogeneity. The finding of this study that both the Classical and Bayesian methods produced similar summary estimates, not only supports the existing belief of the comparability between methods, but also minimizes the concern over any limitation associated with this study's decision to use the Classical methods for random effects analysis.

The third step of *exploring the sources of heterogeneity between individual studies* also revealed several findings. The Breslow-Day test found heterogeneity of studies for the majority of the investigated factors. Although this finding justifies the use of random effects modeling and establishes the basis for the exploration of heterogeneity sources, several researchers have questioned the usefulness of these tests. For example, Oxford researchers applied five commonly used statistical tests on simulated data sets that were truly homogeneous and demonstrated that instead of finding 10% of the simulated meta-analyses as being heterogeneous (as expected from the statistical set up), these tests more often than not over-estimated (e.g. up to 20% of the meta-analyses were considered heterogeneous) or under-estimated (e.g. less than 1% of the meta-analyses were considered heterogeneous) heterogeneity. And when heterogeneity was introduced, these tests could not detect it until the data sets were very heterogeneous.[38]

While the findings of the Oxford investigation demonstrated the importance of using caution in interpreting the results of such heterogeneity tests, others have stated that if found, the exploration of heterogeneity should be the primary focus of the study.[141]

The exploration of heterogeneity in this study using subgroup analyses based upon study design, study quality, and study sample size, revealed that none of the study characteristics were able to "fully explain" the observed heterogeneity. Some researchers have noted similar limitations, in that the usefulness of subgroup analyses are often restricted due to the small number of studies (which can lead to misleading results affected by chance) included in the meta analysis.[142] Although not performed in this study, a regression analysis is considered to be a more sophisticated approach than subgroup analysis for examining the association of treatment effect with other study characteristics. By using the estimate of study results as the dependent variable, and one or more study-level variables as the independent variables (predictors), regression analyses have greater statistical power than subgroup analyses, while also having the ability to simultaneously test multiple characteristics.[142] In the end, the exploration of heterogeneity in this study found that while study design, study quality, and sample size alone did not explain heterogeneity, perhaps a combination of these or other unexplored factors need to be investigated.

Another potential limitation of this study includes the identification of publication bias for 3 of the 18 individual factors (16.7%) (number of prior CD (one versus more than one), previous VD (before prior CD versus none), and augmentation). Although the funnel plots and regression analyses suggest the possibility that the included studies are only a subset of the total body of evidence for these three factors (i.e. publication bias), some investigators have suggested that these methods are flawed and that the observed asymmetry is often due to chance.[143] To demonstrate the lack of reliability of funnel plots, another investigator used the data from 198 Cochrane library meta-analyses to

create two funnel plots for each study, one using sample size and the other using standard error. Even though the same information was used for each of the two funnel plots, it was found that 86% (37 out of 43) of those studies initially found to have an asymmetrical funnel plot (i.e. publication bias) were able to demonstrate a symmetrical funnel plot using the other method.[144] So even though caution must be used in the interpretation of these three individual factors, it appears as if the amount of caution required is questionable.

Implications:

The primary goal of this study was to *educate* those involved with a TOL following prior CD. By addressing the question of *"what factors predict the route of delivery for patients with a prior CD who undergo a trial of labor?"* it was the hope of this study that both those in the clinical and those in the research field would benefit from the findings.

Overall, there are many implications for the findings of this study in the *clinical* field. First of all, both physicians and patients are likely to gain a better understanding of the importance of VBAC through the brief history provided by this study. Secondly, the findings on predictive factors will enable those making the decision of whether or not to attempt a TOL, to make a more tailored and informed decision. Although a way to simultaneously evaluate multiple factors would be ideal and more beneficial to patients and healthcare providers, it was beyond the scope and capabilities of this study. For those who have already decided to undergo a TOL, these findings will either add reassurance to their decision or give them the opportunity to reconsider their options. Lastly, the information from this study may affect the clinical field by leading to

consideration of change in policy or guidelines by professional organizations or administrative bodies, on who should be allowed a TOL following prior CD.

At this time the findings of this study have equally, if not more important implications in the field of *research*. First of all, the systematic review identified the limitation of the existing VBAC literature. Although numerous factors have been proposed, the actual number of studies providing information is unbalanced, completely neglecting several factors which need to be addressed (e.g. demographic and non-clinical factors). Secondly, this study identified the overwhelming lack of "good" or "fair" quality data in the VBAC literature. Again, the main reason why the majority of the studies received a "poor" quality rating is the lack of consideration for confounding. Even though many these studies admitted to the existence of factors with significant associations to TOL outcome, only a small proportion of them decided to account for these in their analysis. This study's attempt to identify those factors associated with TOL outcome, indicate that in fact, there are multiple factors that need to be addressed. Those considering any research in this field need to be aware of these factors and adjust for them accordingly, if they plan on obtaining valid estimates. The third implication of this study is the identification of the need to be more specific with regards to certain factors (e.g. order of previous VD and number of prior CD). While it may be more convenient to only consider these as general categorical factors, this study was able to show the significant differences in estimates between the specific levels of these individual factors (e.g. those with a previous VD before prior CD were significantly less likely to have a VD compared to those with a previous VD after prior CD).

Another possible implication is that the knowledge gained from this study regarding predictive factors could potentially be used to create the ideal screening tool.[145] Compared to existing tools, this screening tool would have increased accuracy and reliability, as well as the ability to provide a larger proportion of the population with useful information. But perhaps more importantly, by simultaneously incorporating these factors, this tool will give the patient and healthcare provider a more useful way of applying research findings. The last implication of this study lies in its contribution to the growing body of evidence in support of meta-analyses of observational studies. Although some researchers such as Shapiro[42] have expressed their concerns regarding the inherent biases and confounding of the observational studies that these studies combine, others[43,44,146,147] have found reason to continue with such research. Along with establishment of the MOOSE guidelines (see Appendix J), the increasing number of meta-analyses of observational studies sends a strong message that researchers are finding ways to use these studies. Diana Petitti summarized this movement best by stating that, "the inability to eliminate bias and confounding does not keep epidemiologists from doing non-experimental studies…it does make the field difficult to work in…but uncertainty, the inability to be definitive, and messiness are not reasons to give up on meta-analysis of non-experimental studies."[44]

For simplicity, we divided the potential implications into the general categories of clinical and research fields. Although easier to follow, this division of fields highlights one of the most important struggles in the field of medicine. Through clinical experience, clinicians are able to generate questions for researchers to answer. While this process seems to work well, the breakdown of *communication* appears to take place for many in

the distribution of answers by researchers to practicing clinicians. Clinicians are often not experienced enough or lack the training to understand the findings of research, making them at times, as uninformed as the patients they treat. For example, how is a physician to interpret and simultaneously consider the summary estimates of all of these individual factors in the context of a patient? Or how is the fractional value (i.e. decimals) of factors such as gravidity, parity, or admission cervical dilation useful, when in clinical practice they are only considered as whole number values. While it is easy to blame the researchers for their lack of preparing data useful to clinicians, one can also place blame on the clinicians for not educating the researchers on what is clinically important or useful. In the end, the usefulness and value of research such as this will depend on the clear and open communication between both clinical and research fields.

Conclusion:

In conclusion, this study has been able to identify several factors that are significantly associated with TOL outcome in those with a prior CD. In the setting of rising CD rates and the need for more precise methods of selecting those individual for TOL, this study has attempted to educate patients, physicians, and researchers. In doing so, it is the hope of this study that more informed decisions would be made and that an effort towards open communication between the clinical and research fields will be attempted. Hopefully this study will act as another stepping stone toward the goal of one day being able to respond to the question of "predicting vaginal birth after CD success: is it possible?" with a confident "yes".

REFERENCES

[1] Zinberg S. Vaginal Delivery After Previous Cesarean Delivery: A Continuing Controversy. Clin Obstet & Gynecol. 2001; 44(3): 561-570.

[2] Speert H. Obstetrics and Gynecology in America: A History. Baltimore, MD: Waverly Press, 1980: 153-156.

[3] Cragin EB. Conservatism in obstetrics. N Y Med J. 1916; 104: 1-3.

[4] Chua S, Arulkumaran S. Trial of scar. Aust N Z Obstet Gynaecol. 1997; 36:6.

[5] Marieskind HI. Cesarean Section in the United States: Has it changed since 1979? Birth. 1989; 16(4): 196-202.

[6] Shiono PH et al. Reasons for the Rising Cesarean Delivery Rates: 1978-1984. Obstet & Gynecol. 1987; 69(5): 696-700.

[7] U.S. Department of Health and Human Services: Public Health Service, National Institutes of Health: Repeat Cesarean Birth. IN Cesarean Childbirth. Washington, DC: U.S. Government Printing Office, 1991; 351-371.

[8] ACOG Committee on Obstetrics: Maternal and Fetal Medicine. Statement on vaginal birth after cesarean (VBAC). Washington DC: American College of Obstetricians and Gynecologists, 1985.

[9] Flamm BL et al. Vaginal Birth After Cesarean Delivery: Results of a 5-Year Multicenter Collaborative Study. Obstet Gynecol. 1990; 76(5): 750-754.

[10] Phelan JP et al. Vaginal birth after cesarean. Am J Obstet Gynecol. 1987; 157: 1510-1515.

[11] Ventura SJ et al. Advance report of final natality statistics, 1993. National Center for Health Statistics. Monthly Vital Stat Rep 1993; 44: 72.

[12] Flamm BL. Once a Cesarean, Always a Controversy. Obstet Gynecol. 1997; 90(2): 312-315.

[13] Miller DA et al. Vaginal Birth After Cesarean: A 10-Year Experience. Obstet Gynecol. 1994; 84(2): 255-258.

[14] Jones RO et al. Rupture of low transverse cesarean scars during trial of labor. Obstet Gynecol. 1991; 77: 815-817.

[15] Scott JR. Mandatory trial of labor after cesarean delivery: An alternative viewpoint. Obstet Gynecol. 1991; 77: 811-814.

[16] Pitkin. Once a Cesarean? Obstet Gynecol. 1991; 77(6): 939.

[17] McMahon MJ et al. Comparison of a Trial of Labor with an Elective Second Cesarean Section. N Engl J Med. 1996; 335(10): 689-695.

[18] Rosen MG, Dickinson JC, Westhoff CL. Vaginal Birth After Cesarean: A Meta-Analysis of Morbidity and Mortality. Obstet Gynecol. 1991; 77(3): 465-470.

[19] Lavin JR et al. Vaginal Delivery in Patients with Prior Cesarean Section. Obstet Gynecol. 1982; 59(2): 135-148.

[20] Weinstein D et al. Vaginal birth after cesarean section: current opinion. Int J of Gynecol Obstet. 1996; 53: 1-10.

[21] McMahon MJ. Vaginal Birth After Cesarean. Clin Obstet Gynecol. 1998; 41(2): 369-381.

[22] Rosen MG, Dickinson JC. Vaginal Birth After Cesarean: A Meta-Analysis of Indicators for Success. Obstet Gynecol. 1990; 76: 865-869.

[23] Troyer LR, Parisi VM. Obstetric parameters affecting success in a trial of labor: Designation of a scoring system. Am J Obstet Gynecol. 1992; 167(4): 1099-1103.

[24] Flamm BL, Geiger AM. Vaginal Birth After Cesarean Delivery: An Admission Scoring System. Obstet Gynecol. 1997; 90(6): 907-910.

[25] Vinueza CA et al. Predicting the Success of a Trial of Labor with a Simple Scoring System. J Reprod Med. 2000; 45(4): 332-336.

[26] Weinstein D et al. Predictive score for vaginal birth after cesarean section. Am J Obstet Gynecol. 1996; 174(1): 192-198.

[27] Whiteside DC et al. Factors associated with successful vaginal delivery after cesarean section. J Reprod Med. 1983; 28: 785-788.

[28] Pickhardt MG et al. Vaginal birth after cesarean delivery: Are there useful and valid predictors of success or failure? Am J Obstet Gynecol. 1992; 166(6): 1811-1815.

[29] Ananonymous. Trends in the Attendant, Place, and Timing of Births, and in the Use of Obstetric interventions: United States, 1989-1997. National Vital Statistics Reports 1999; 47(27): 1-16.

[30] American College of Obstetricians and Gynecologists: Vaginal Birth After Previous Cesarean Delivery. Washington, DC, American College of Obstetricians and Gynecologists, 1999. ACOG Practice Bulletin #5

[31] Chung A et al. Cost-Effectiveness of a Trial of Labor After Previous Cesarean. Obstet Gynecol. 201; 97: 932-941.

[32] Leung AS et al. Uterine rupture after previous cesarean delivery: Maternal and fetal consequences. Am J Obstet Gynecol. 1993; 169(4): 945-950.

[33] CD Summary – Oregon Department of Human Services: Cesarean Births in Oregon. 6/2002; 51(12).

[34] Gregoire G et al. Selecting the language of the publications included in a meta-analysis: is there a tower of babel bias? J Clin Epidemiol. 48(1): 159-163. 1995.

[35] ACOG Committee Opinion. Mode of Term Singleton Breech Delivery. 2001; 265.

[36] Hannah ME et al. Planned cesarean section versus planned vaginal birth for breech presentation at term: a randomized multicenter trial. Term Breech Trial Collaborative Group. Lancet. 2000; 356: 1375-83.

[37] DerSimonian R, Laird N. Meta-analysis in Clinical Trials. Controlled Clinical Trials. 1986; 7: 177-188.

[38] Gavaghan DJ et al. An evaluation of homogeneity tests in meta-analysis in pain using simulations of patient data. Pain. 2000; 85: 415-24.

[39] Vine SK et al. Fitting Bayesian multiple random effects models. Statistic Computing. 1996; 6: 337-46.

[40] Gilks WR et al. A language and program for complex Bayesian modeling. The Statistician. 1994; 43: 169-78.

[41] Smith TC et al. Bayesian approaches to random effects meta-analysis: a comparative study. Statistics Medicine. 1995; 14: 2685-2699.

[42] Shapiro S. Meta-analysis/Shmeta-analysis. Am J Epidemiol. 1994; 140(9): 771-778.

[43] Greenland S. Can Meta-analysis Be Salvaged. Am J Epidemol. 1994; 140(9): 783-791.

[44] Petitti DB. Of Babies and Bathwater. Am J Epidemiol. 1994; 140(9): 779-782.

[45] Sutton AJ et al. Methods of Meta-Analysis in Medical Research. John Wiley & Sons, Ltd. New York, New York. 2000.

[46] Abitbol MM. The Cephalopelvic Disproportion Index: combined fetal sonography and X-ray pelvimetry for early detection of cephalopelvic disproportion. J of Repro Med. 1991; 36(5): 369-73.

[47] Adair, CD et al. Labor Induction in Patients with Previous Cesarean Section. Am J of Perinatology. 1995; 12(6): 450-454.

[48] Amir W et al. Trial of Labor without Oxytocin in Patients with a Previous Cesarean Section. Am J of Perinatology. 1987; 4(2): 140-143.

[49] Arulkumaran S et al. Oxytocin augmentation in dysfunctional labour after previous caesarean section. Br J of Obstet Gynecol. 1989; 96: 939-941.

[50] Asakura H, Myers SA. More than one previous cesarean delivery: A 5-year experience with 435 patients. Obstet Gynecol. 1995; 85: 924-929.

[51] Bias JMJ, et al. Vaginal birth after cesarean section in a population with a low overall cesarean section rate. Eur J of Obstet Gynecol. 2001; 96: 158-162.

[52] Beall M et al. Vaginal delivery after cesarean section in women with unknown types of uterine scar. J of Repro Med. 1984; 29(1): 31-35.

[53] Bedoya C et al. A trial of labor after cesarean section in patients with or without a prior vaginal delivery. Int J Gynecol Obstet. 1992; 39: 285-289.

[54] Blackwell SC et al. Vaginal birth after cesarean in the diabetic gravida. J of Repro Med. 2000; 45: 987-990.

[55] Blanchette H et al. Is vaginal birth after cesarean safe? Experience at a community hospital. Am J Obstet Gynecol. 2001; 184: 1478-87.

[56] Brettel F et al. Vaginal birth following two previous cesarean sections. Eur J of Obstet Gynecol. 2001; 94: 23-26.

[57] Callahan C et al. Safety and efficacy of attempted vaginal birth after cesarean beyond the estimated date of delivery. J of Repro Med. 1999; 44: 606-610.

[58] Carroll SG et al. Management of antepartum spontaneous rupture after one previous cesarean section. Eruo J of Obstet Gynecol. 1990; 34: 173-178.

[59] Caughey AB et al. Trial of labor after cesarean delivery: the effect of previous vaginal delivery. Am J Obstet Gynecol. 1998; 179(4): 938-941.

[60] Chang et al. An analysis of 121 trial labors following a previous cesarean section. Changgeng Yi Xue Za Zhi. 1987; 10(1): 28-36.

[61] Chauhan SP et al. Mode of delivery for the morbidly obsess with prior cesarean delivery: vaginal versus repeat cesarean section. Am J Obstet Gynecol. 2001; 185: 349-54.

[62] Chua S et al. Trial of labour after previous cesarean section: obstetric outcome. Aust NZ J Obstet Gynecol. 1989; 29: 12-17.

[63] Clark SL et al. Effect of Indication for previous cesarean section on subsequent delivery outcome in patients undergoing a trial of labor. J of Repro Med. 1984; 29(1): 22-25.

[64] Coleman TL et al. Vaginal birth after cesarean among women with gestational diabetes. Am J Obstet Gynecol. 2001; 184: 1104-7.

[65] Cowan RK et al. Trial of labor following cesarean delivery. Obstet Gynecol. 1994; 83: 933-6.

[66] Duff P et al. Outcome of trial of labor in patients with a single previous low transverse cesarean section for dystocia. Obstet Gynecol. 1988; 71: 380-384.

[67] Edelin KC. The effect of induced abortion on the risks and outcome of a trial of labor after a previous cesarean birth. J of Nat Med Assoc. 1988; 80(9): 993-5.

[68] Eglinton GS et al. Outcome of a trial of labor after prior cesarean delivery. J of Repro Med. 1984; 29(1): 3-8.

[69] Eriksen NL, Buttino L. Vaginal birth after cesarean: a comparison of maternal and neonatal morbidity to elective repeat cesarean section. Am J of Perinatology. 1989; 6(4): 375-9.

[70] Farmakides G et al. Vaginal birth after two or more previous cesarean sections. Am J Obstet Gynecol. 1987; 156: 565-6.

[71] Flamm BL et al. Oxytocin during labor after previous cesarean section: results of a multicenter study. Obstet Gynecol. 1987; 70: 709-712.

[72] Flamm BL et al. Vaginal birth after cesarean section: results of a multicenter study. AM J Obstet Gynecol. 1988; 158: 1079-84.

[73] Flamm BL, Goings JR. Vaginal birth after cesarean section: is suspected fetal macrosomia a contraindication? Obstet Gynecol. 1989; 74: 694-97.

[74] Flamm BL et al. External cephalic version after previous cesarean section Am J Obstet Gynecol. 1991; 165: 370-2.

[75] Flamm BL et al. Elective repeat cesarean delivery versus trial of labor: a prospective multicenter study. Obstet Gynecol. 1994; 83: 927-32.

[76] Fraser W et al. Randomized controlled trial of a prenatal vaginal birth after cesarean section education and support program. Am J Obstet Gynecol. 1997; 176(2): 419-25.

[77] Goldman G et al. Effects of patient, physician and hospital characteristics on the likelihood of vaginal birth after previous cesarean section in Quebec. Can Med Assoc J. 1990; 143(10): 1017-24.

78 Granovsky-Grisaru S et al. The management of labor in women with more than one uterine scar: is a repeat cesarean section really the only "safe" option? J Perinat Med. 1994; 22: 13-17.

[79] Hadley CB. An evaluation of the relative risks of a trial of labor versus elective repeat cesarean section. Am J of Perinatology. 1986; 3(2): 107-114.

[80] Hangsleben KL. VBAC program in a nurse-midwifery service: five years of experience. J of Nurse-Midwifery. 1989; 34(4): 179-184.

[81] Hansell RS et al. Vaginal birth after two or more cesarean sections: a five-year experience. Birth. 1990; 17(3): 146-151.

[82] Holt VL, Mueller BA. Attempt and success rates for vaginal birth after cesarean section in relation to complications of the previous pregnancy. Ped and Perinate Epid. 1997; 11(supp): 63-72.

[83] Horenstein JM et al. Oxytocin use during a trial of labor in patients with previous cesarean section. J of Repro Med. 1984; 29(1):26-30.

[84] Horensten JM, Phelan JP. Previous cesarean section: the risks and benefits of oxytocin usage in a trial of labor. Am J Obstet Gynecol. 1985; 151: 564-9.

[85] Hoskins IA, Gomez JL. Correlation between maximum cervical dilation at cesarean delivery and subsequent vaginal birth after cesarean delivery. Obstet Gynecol. 1997; 89: 591-3.

[86] Huang WH et al. Interdelivery interval and the success of vaginal birth after cesarean delivery. Obstet Gynecol. 2002; 99: 41-44.

[87] Jakobi P et al. Evaluation of prognostic factors for vaginal delivery after cesarean section. J of Repro Med. 1993; 38(9): 729-33.

[88] Jongen VHWM et al. Vaginal delivery after previous cesarean section for failure of second state of labor. Br J Obstet Gynecol. 1998; 105: 1079-81.

[89] King DE, Lahin K. Socioeconomic factors and the odds of vaginal birth after cesarean delivery. JAMA. 1994; 272: 524-9.

[90] Lai SF, Sidek S. Delivery after a lower segment cesarean section. Singapore Med J. 1993; 34: 62-66.

[91] Lao TT et al. Is x-ray pelvimetry useful in a trial of labor after cesarean section? Eur J Obstet Gynecol. 1987; 24: 277-83.

[92] Learman LA et al. Predictors of repeat cesarean delivery after trial of labor: do any exist? J Am Coll Surg. 1996; 182: 257-62.

[93] Lelaidier C et al. Mifepristone for labor induction after previous cesarean section. Br J Obstet Gynecol. 1994; 101: 501-3.

[94] Lovell R. Vaginal delivery after cesarean section: factors influencing success rates. Aust NZ J Obstet Gynecol. 1996; 36(1): 4-8.

[95] Macones GA et al. Predicting outcomes of trials of labor in women attempting vaginal birth after cesarean delivery: a comparison of multivariate methods with neural networks. Am J Obstet Gynecol. 2001; 184: 409-13.

[96] Mahmood TA, Grant JM. The role of radiological pelvimetry in the management of patients who have had a previous cesarean section. J Obstet Gynecol. 1987; 8: 24-8.

[97] McClain CS. Patient decision making: the case of delivery method after a previous cesarean section. Cult, Med and Psych. 1987; 11: 494-508.

[98] McNally OM. Induction of labor after 1 previous cesarean section. Aust NZ J Obstet Gynecol. 1999; 39(4): 425-429.

[99] Meier PR, Porreco RP. Trial of labor following cesarean section: a two-year experience. Am J Obstet Gynecol. 1982; 144: 671-7.

[100] Miller DA et al. Vaginal birth after cesarean: a 10-year experience. Obstet Gynecol. 1994; 84: 255-8.

[101] Miller DA et al. Vaginal birth after cesarean section in twin gestation. Am J Obstet Gynecol. 1996; 175: 194-8.

[102] Miller M, Leader LR. Vaginal delivery after cesarean section. Aust NZ J Obstet Gynecol. 1992; 32(3): 213-6.

[103] Mor-Yosef S et al. Vaginal delivery following one previous cesarean birth: nation wide survey. Asia-oceania J Obstet Gynecol. 1990; 16(1): 33-7.

[104] Morgan MA, Thurnau GR. Efficacy of the fetal-pelvic index for delivery of neonates weighing 4000 grams or greater: a preliminary report. Am J Obstet Gynecol. 1988; 158: 1133-7.

[105] Nguyen TV et al. Vaginal birth after cesarean section at the University of Texas. J of Repro Med. 1992; 37(10): 880-2.

[106] Novas J et al. Obstetric outcome of patients with more than one previous cesarean section. Am J Obstet Gynecol. 1989; 160: 364-7.

[107] Ollendorff DA et al. Vaginal birth after cesarean section for arrest of labor: is success determined by maximum cervical dilation during prior labor? Am J Obstet Gynecol. 1988; 159: 636-9.

[108] Paterson CM, Saunders NJ. Mode of delivery after one cesarean section: audit of current practice in a health region. BMJ. 1991; 303: 818-21.

[109] Paul RH et al. Trial of labor in the patient with a prior cesarean birth. Am J Obstet Gynecol. 1985; 151: 297-304.

[110] Phelan JP et al. Previous cesarean birth: Trial of labor in women with macrosomic infants. J of Rep Med. 1984; 29(1): 36-40.

[111] Phelan JP et al. Twice a cesarean, always a cesarean? Obstet Gynecol. 1989; 73: 161-5.

[112] Porreco RP, Meier PR. Trial of labor in patients with multiple previous cesarean sections. J of Repro Med. 1983; 28(11): 770-2.

[113] Pruett KM et al. Is vaginal birth after two or more cesarean sections safe? Obstet Gynecol. 1988; 72: 163-5.

[114] Rageth JC et al. Delivery after previous cesarean: a risk evaluation. Obstet Gynecol. 1999; 93: 332-7.

[115] Rudick V et al. Epidural anesthesia for planned vaginal delivery following previous cesarean section. Obstet Gynecol. 1984; 64: 621-3.

[116] Sakala EP et al. Oxytocin use after previous cesarean: why a higher route of failed labor trial? Obstet Gynecol. 1990; 75: 356-9.

[117] Sakala EP et al. Epidural anesthesia: effect on the likelihood of a successful trial of labor after cesarean section. J of Repro Med. 1990; 35(9): 886-90.

[118] Schachter M et al. External cephalic version after previous cesarean section – a clinical dilemma. Int J Gynecol Obstet. 1994; 45: 17-20.

[119] Shipp TD et al. Labor after previous cesarean: influence of prior indication and parity. Obstet Gynecol. 2000; 95: 913-6.

[120] Silver RK, Gibbs RS. Predictors of vaginal delivery in patients with a previous cesarean section, who require oxytocin. Am J Obstet Gynecol. 1987; 156: 57-60.

[121] Sims EJ et al. Vaginal birth after cesarean: to induce or not to induce. Am J Obstet Gynecol. 2001; 184: 1122-4.

[122] Socol ML, Peaceman AM. Vaginal birth after cesarean: an appraisal of fetal risk. Obstet Gynecol. 1999; 93(5): 674-9.

[123] Spaans WA et al. Risk factors at cesarean section and failure of subsequent trial of labor. Eur J Obstet Gynecol. 2002; 100: 163-6.

[124] Stafford RS. The impact of nonclinical factors on repeat cesarean section. JAMA. 1991; 265: 59-63.

[125] Stovall TG et al. Trial of labor in previous cesarean section patients, excluding classical cesarean sections. Obstet Gynecol. 1987; 70: 713-7.

[126] Strong TH et al. Vaginal birth after cesarean delivery in the twin gestation. Am J Obstet Gynecol. 1989; 161: 29-32.

[127] Stronge JM, McQuillan K. Factors affecting mode of delivery in labor following a signle previous birth by cesarean section. J Obstet Gynecol. 1996; 16(5): 353.

[128] Thubisi M et al. Vaginal delivery after previous cesarean section: is x-ray pelvimetry necessary? Br J Obstet Gynecol. 1993; 100: 421-4.

[129] Thurnau GR et al. The fetal-pelvic index: a method of identifying fetal-pelvic disproportion in women attempting vaginal birth after previous cesarean delivery. Am J Obstet Gynecol. 1991; 165: 353-8.

[130] Van Gelderen CJ et al. Labor in patients with a cesarean section scar. SAMJ. 1986; 70: 529-532.

[131] Walton DL et al. Vaginal birth after cesarean section: acceptance and outcome at a rural hospital. J of Repro Med. 1993; 38(9): 716-8.

[132] Wax JR et al. Twin vaginal birth after cesarean. Connecticut Med. 2000; 64(4): 205-208.

[133] Wright M. The value of X-ray pelvimetry in previous cesarean section pregnancies. SAMT. 1985; 68: 409-411.

[134] Yasumizu T et al. Trial of labor in the patient with a prior cesarean section. Yamanashi Med J. 1993; 8(3): 113-9.

[135] Yasumizu T et al. Trial of vaginal birth following cesarean section for arrest disorders of labor: analysis of patients with well-documented medical records. Asia-oceania J Obstet Gynecol. 1994; 20(4): 407-13.

[136] Yetman TJ, Nolan TE. Vaginal birth after cesarean section: a reappraisal of risk. Am J Obstet Gynecol. 1989; 161: 1119-23.

[137] Zelop CM et al. Trial of labor after 40 weeks' gestation in women with prior cesarean. Obstet Gynecol. 2001; 97: 391-3.

[138] Zelop CM et al. Outcomes of trial of labor following previous cesarean delivery among women with fetuses weighing >4000g. Am J Obstet Gynecol. 2001; 185(4): 903-5.

[139] Zorlu CG et al. Vaginal birth following unmonitored labor in patients with prior cesarean section. Gynecol Obsete Invest. 1996; 42: 222-26.

[140] Coltart TM et al. Outcome of a second pregnancy after a previous elective cesarean section. Br J Obstet Gynecol. 1990; 97(12): 1140-3.

[141] DerSimonian R. Resolving Discrepancies between a Meta-analysis and a subsequent large controlled trial. JAMA. 1999; 282: 664-670.

[142] Song F et al. Methods for exploring heterogeneity in meta-analysis. Evaluation & The Health Professions. 2001; 24(2): 126-151.

[143] Terrin NC et al. Publication bias, chance, and heterogeneity: how researchers interpret the funnel plot. BMC Meeting Abstracts: 9th International Cochrane Colloquium. 2001.

[144] Tang J, Liu JL. Misleading funnel plot for detection of bias in meta-analysis. Journal of Clinical Epidemiology. 2000; 53: 477-84.

[145] Guise JM et al. Vaginal Birth After Cesarean. Agency for Healthcare Research and Quality. Contract Number 290-97-0018, Task Order No.9.

[146] Concato J et al. Randomized, Controlled Trials, Observational Studies and the Hierarchy of Research Designs. N Engl J Med. 2000; 342(25): 1887-1892.

[147] Benson K, Hartz AJ. A Comparison of Observational Studies and Randomized, Controlled Trials. N Engl J Med. 2000; 342(25): 1878-1886.

[148] Harris RP, Helfand M, Woolf SH, et al. Methods of the third U.S. Preventive Services Task Force. Am J Prev Med. 2001; 20(3S): 21-35.

[149] Undertaking Systematic Reviews of Research Effectiveness: CRD's Guidance for those Carrying Out or Commissioning Reviews. CRD Report Number 4 (2nd ed). NHS centre for Reviews and Dissemination; York, England. March 2001.

Appendix A: Database Search Parameters

Databases: MEDLINE (1966 to 2002), HealthSTAR (1975 to 2002)

1. Vaginal birth after cesarean/ or "vaginal birth after cesarean".mp.
2. (trial of labor or trial of labour or trial of scar$).mp.
3. Delivery/ or Episiotomy/ or Extraction, obstetrical/ or Home childbirth/ or Labor, induced/ or Natural childbirth/ or Version, fetal/
4. (vaginal birth or vaginal delivery or uterine rupture).mp. [mp=title, abstract, registry number word, mesh subject heading]
5. exp Labor/
6. 2 or 3 or 4 or 5
7. exp cesarean section/ or "cesarean".mp.
8. 6 and 7
9. 1 or 8
10. limit 9 to human
11. limit 10 to English language
12. 10 not 11
13. limit 12 to abstracts
14. 11 or 13
15. exp risk assessment/ or "risk assessment".mp.
16. exp probability/ or "probability".mp.
17. Predictive value of tests/
18. previous vaginal delivery.mp.
19. Gestational age/ or "gestational age".mp.
20. "SPONTANEOUS LABOR".mp.
21. Birth weight/ or "birth weight".mp.
22. Fetal weight/ or "fetal weight".mp.
23. exp labor presentation/ or Oxytocin/ or "cervical dilation".mp.
24. exp treatment outcome/ or Pregnancy outcome/ or "outcome".mp.
25. Cesarean section, repeat/ or "repeat cesarean".mp.
26. 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25
27. 14 and 26

Appendix B: Developed Countries (DCs)

CIA World Factbook 2001

The top group in the hierarchy of developed countries (DCs), former USSR/Easter Europe (former USSR/EE), and less developed countries (LDCs); includes the market-oriented economies of the mainly democratic nations in the Organization for Economic Cooperation and Development (OECD), Bermuda, Israel, South Africa, and the European ministates; also known as the First World, high-income countries, the North, industrial countries; generally have a per capita GDP in excess of $10,000 although four OECD countries and South Africa have figures well under $10,000 and two of the excluded OPEC countries have figures of more than $10,000; the 35 DC are: *Andorra, Australia, Austria, Belgium, Bermuda, Canada, Denmark, Faroe Islands, Finland, France, Germany, Greece, Holy See, Iceland, Ireland, Israel, Italy, Japan, Liechtenstein, Luxembourg, Malta, Mexico, Monaco, Netherlands, NZ, Norway, Portugal, San Marino, South Africa, Spain, Sweden, Switzerland, Turkey, UK, US*; note – similar to the new International Monetary Fund (IMF) term "advanced economies" which adds *Hong Kong, South Korea, Singapore*, and *Taiwan* but drops Malta, Mexico, South Africa, and Turkey.

Appendix C: WINBUGS® Programming Codes

---------- CATEGORICAL DATA SET UP ----------

USING NON-INFORMATIVE priors (d)

```
model
{
  for( i in 1 : 12* ) {
        y.no[i] ~ dbin(pno[i], n.no[i])
        y.yes[i] ~ dbin(pyes[i], n.yes[i])
        logit(pno[i]) <- mu[i]
        logit(pyes[i]) <- mu[i] + delta[i]
        mu[i] ~ dnorm(0.0,1.0E-5)
        delta[i] ~ dnorm(d, tau)
  }
  d ~ dnorm(0.0,1.0E-6)
  tau ~ dgamma(0.001,0.001)
  sigma <- 1 / sqrt(tau)
  or <- exp(d)
}

inits
list(d = 0, tau = 1, mu = c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0*),
delta = c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0*))
```

y = vbac
n = total # in group with/without variable
yes = yes to variable
no = no to variable
y.yes [ ] n.yes [ ] y.no[ ] n.no[ ]

---------- CONTINOUS DATA FORMAT ----------

USING NON-INFORMATIVE priorS (mu.vbac, mu.ftol)

```
model
{
  for( i in 1 : 3* ) {
  y.vbac[i] ~ dnorm(mu.vbac, tau.vbac[i])
  y.ftol[i] ~ dnorm(mu.ftol, tau.ftol[i])
  tau.vbac[i] <- 1/sd.vbac[i]*sd.vbac[i]
  tau.ftol[i] <- 1/sd.ftol[i]*sd.ftol[i]
  }
 mu.vbac ~ dnorm(0.0,1.0E-6)
 mu.ftol ~ dnorm(0.0,1.0E-6)
  diff <- mu.vbac - mu.ftol
 pval <- step(diff)
}

init
list(mu.vbac = 0, mu.ftol = 0)
```

y.vbac [ ] sd.vbac [ ] y.ftol [ ] sd.ftol [ ]

**Italicized* numbers are variable and specific to the number of studies included in the analysis.

Appendix D: Studies Excluded at the Full-Text Level

Reason for Study Exclusion (n)	Author - Year
No data for the topic/focus on uterine rupture (58)	Arulkumaran – 1989, Arulkumaran – 1992, Bartha – 2000, Beckley – 91, Caughey – 1999, Chapman – 1997, Chazotte – 1990, Clark – 1988, Davies – 1996, Esposito – 2000, Fedorkow, Ferguson – 1998, French – 1996, Geraldine – 2001, Goetzl – 2001, Gregory (NR), Gregory – 1994, Gregory – 1999, Grubb – 1996, Huang – 2002, Hueston – 1994, Kline – 1993, Lelaidier – 1994, Leung – 1993, Lipson – 1984, Lonky – 1989, Lynch – 1996, Lyndon-Rochelle – 2001, McClain – 1985, McClain – 1990, Michaels – 1988, Murphy – 1989, National Vital Stats – 2001, Nielsen – 1984, Nielsen – 1985, Norman – 1993, O'Sullivan – 1981, Pel – 1995, Phelan – 1998, Plaut – 1999, Pruett – 1988, Ramin – 1992, Ravasia – 2000, Rozenberg – 1996, Rozenberg – 1999, Shimonovitz – 2000, Shiono – 1987, Shipp – 2001, Sieck – 1997, Socol –1993, Tanik – 1996, Tucker – 1993, Veridiano – 1989, Wing – 1998, Wittich – 2000, Zelop – 1999, Zelop – 2000,
Review/letter/commentary (34)	Bhal – (NR), ACOG – 1996, ACOG – 2001, Beckett – 2001, Boyers – 1998, Catanzarite - NR, Chang – 1997, Cohen – NR, Daviss – 2001, Flamm - 1992, Flamm – 2001, Fribourg – 1987, Fuller Miller – 1985, Hobbs – 1994, Kobelin – 2001, Lavin – 1982, MacDonald – (NR), Macones – 1999, Maouris – 1987, Marieskind – 1989, McMahon – 1998, Meehan – 1988, Ophir – 1988, Pridjian – 1992, Rosen – 1990, Rosen – 1990, SOGC guidelines – 97, Sutcliffe – 1994, Turner – 1997, Waldman – 2001, Walker – 2002, Weinstein – 1995, Zinberg – 2001,
Study population include those recruited prior to 1980 (20)	Benedetti – 1982, Chemlow – 1992, Demianczuk – 1982, Flamm – 1984, Gellman – 1983, Gibbs – 1980, Impey – 1988, Jarrell – 1985, Krishnamurthey – 1991, Mahmood – 1989, Meehan – 1988, Molloy – 1987, Mootbar – 1984, Ngu – 1985, Odeh – 1997, Prendergast – 1985, Seitchik – 1982, Targett – 1988, Whiteside – 1983, Yeh - 1984
Study conducted in a non-developed country (13)	Abu-Ghazzeh – 2000, Chattophadhyay – 1988, Chattophadhyay – 1994, Chi – 1983, Dhall – 1987, Dyack – 1997, Ojo – 1989, Perveen – 1997, Sakka – 1998, Singh – 1986, Van der Walt – 1994, Wadhawan – 1983, Yamani - 1999
Study population does not include those with a prior cesarean delivery (9)	Berkowitz – 1989, Fletcher – 1998, Jagani – 1981, Morgan – 1986, Sandmire – 1993, Schussman – 1982, Westgate – 1994, Witter – 1992, Zakut - 1981
Case reports/series of <10 subjects (5)	Fogarty – 1993, Lau – 1994, McKenna – 1988, Raskin – 1999, Uppington – 1983,
Incorrect comparison groups employed (4)	Del Valle – 1994, Goldman – 1990, King – 1994, Stafford - 1991
Vertical incision (3)	Adair – 1996, Halperin – 1988, Naef -1995
Breech TOL (2)	Ophir – 1989, Sarno - 1989
Error in data (1)	Wagner - 1999

NR = not reported

Appendix E: Quality Ratings by Study Design

RCT

Study, year	Random assignment?	Groups similar at baseline?	Intention-to-treat analysis?	Differential loss to follow-up or overall high loss to follow-up?	Score
Thubisi, 93	Y	Y	Y	NA	GOOD
Fraser, 97	Y	Y	N	Y/N	FAIR

COHORT – PROSPECTIVE

Author, Year	Comparable groups assembled/ Database representative for study	Clear definition of comparison groups/sufficient description of distribution of prognostic factors	Consider/Adjust for potential important confounders	Score
Abitbol, 91	N	N	N	POOR
Bais, 2000	Y	N	N	POOR
Blanchette, 01	Y	N	N	POOR
Chang, 87	Y	N	N	POOR
Chauhan, 2001	Y	N	Y	POOR
Cowan, 94	Y	N	N	POOR
Duff, 88	Y	N	N	POOR
Flamm, 87	Y	N	N	POOR
Flamm, 88	Y	N	N	POOR
Flamm, 89	Y	N	N	POOR
Flamm, 90	Y	N	N	POOR
Flamm, 94	Y	N	N	POOR
Flamm, 97	Y	Y/N	Y	GOOD
Granovsky, 94	N	NA	N	POOR
Horenstein, 84	Y	N	N	POOR
Horenstein, 85	Y	N	N	POOR
McClain, 87	Y	N	N	POOR
Miller DA, 94	Y	N	N	POOR
Miller DA, 96	N	NA	N	POOR
Miller M, 92	Y	N	N	POOR
Morgan, 88	Y	N	N	POOR
Phelan, 87	Y	N	N	POOR
Phelan, 89	Y	N	N	POOR
Rudick, 84	N	NA	N	POOR
Silver, 87	Y	N	N	POOR
Sims, 01	Y	N	N	POOR
Stovall, 87	Y	N	N	POOR
Stronge, 1996	Y	N	Y/N	FAIR
Thurnau, 91	Y	N	N	POOR
Wright, 85	Y	N	N	POOR
Yasumizu, 94	Y	N	N	POOR

APPENDIX E: Quality Ratings by Study Design – Continued

COHORT - RETROSPECTIVE

Author, Year	Comparable groups assembled/ Database representative for study	Clear definition of comparison groups/sufficient description of distribution of prognostic factors	Consider/Adjust for potential important confounders	Score
Adair, 95	N	NA	N	POOR
Amir, 87	Y	N	N	POOR
Arulkumaran, 89	Y	N	N	POOR
Asakura, 95	Y	Y	N	POOR
Beall, 84	Y	Y	N	POOR
Bedoya, 92	Y	N	N	POOR
Blackwell, 2000	N	N	N	POOR
Brettelle, 2001	Y	N	N	POOR
Callahan, 99	Y	N	N	POOR
Carroll, 89	Y	N	N	POOR
Caughey, 98	Y	Y	Y	GOOD
Chua, 89	Y	N	N	POOR
Clark, 84	Y	N	N	POOR
Coleman, 2001	Y	Y	N	POOR
Coltart, 90	Y	N	N	POOR
Edelin, 88	Y	N	N	POOR
Eglinton, 84	Y	N	N	POOR
Eriksen, 1989	N	NA	N	POOR
Farmakides, 87	Y	N	N	POOR
Hadley, 86	Y	N	N	POOR
Hangsleben, 89	Y	N	N	POOR
Hansell, 90	Y	N	N	POOR
Holt, 97	Y	N	N	POOR
Hoskins, 97	Y	N	N	POOR
Huang, 2002	Y	Y	Y	GOOD
Jakobi, 93	Y	N	Y	FAIR
Jongen, 98	Y	N	N	POOR
Lai, 93	Y	N	N	POOR
Lao, 87	Y	N	N	POOR
Mahmood, 87	Y	N	N	POOR
McMahon, 96	Y	Y	N	POOR
McNally, 99	Y	N	Y	FAIR
Meier, 82	Y	N	N	POOR
Mor-Yosef, 90	N	NA	N	POOR
Nguyen, 92	Y	N	N	POOR
Novas, 89	Y	N	N	POOR
Ollendorf, 88	Y	N	N	POOR
Paterson, 91	Y	N	N	POOR
Paul, 85	Y	N	N	POOR
Phelan, 84	Y	N	N	POOR

APPENDIX E: Quality Ratings by Study Design – Continued

COHORT – RETROSPECTIVE - Continued

Author, Year	Comparable groups assembled/ Database representative for study	Clear definition of comparison groups/sufficient description of distribution of prognostic factors	Consider/Adjust for potential important confounders	Score
Porreco, 83	Y	N	N	POOR
Pruett, 88	Y	N	N	POOR
Rageth, 99	Y*	N	N	POOR
Sakala, 90 (epi)	Y	Y	N	POOR
Sakala, 90 (oxy)	Y	N	N	POOR
Shipp, 2000	Y*	N	N	POOR
Socol, 99	Y	N	N	POOR
Spaans, 2002	Y	N	N	POOR
Strong, 89	N*	NA	N	POOR
Troyer, 92	Y	N	N	FAIR
VanGelderen, 86	Y	N	N	POOR
Vinueza, 2000	Y	Y	N	FAIR
Walton, 93	Y	N	N	POOR
Wax, 99	N*	NA	N	POOR
Weinstein, 96	Y	N	Y	FAIR
Yasumizu, 93	Y	N	N	POOR
Yetmen, 89	Y*	N	N	POOR
Zelop, 2001 (GA)	Y	Y/N	Y	FAIR
Zelop, 2001(mac)	Y	Y/N	Y	FAIR
Zorlu, 96	Y	N	N	POOR

CASE-CONTROL

Author, Year	Case definition explicit	Nonbiased selection of cases/controls - *(Controls randomly selected)*	Measurement of exposure accurate and applied equally	Appropriate attention to confounders	Score-
Learman, 96	Y	N	Y/N	N	POOR
Lovell, 96	Y	N	Y/N	N	POOR
Macones, 2001	Y	N	Y/N	Y	FAIR
Pickhardt, 92	Y	N	Y/N	Y	FAIR

CASE-SERIES

Author, Year	Representative sample selected from a relevant population	Individuals entered the survey at a similar point in their disease progression	If comparison of sub-series, sufficient description of the series and distribution of prognostic factors	Score
de Meeus, 98	Y	Y/N	N	FAIR
Flamm, 91	Y	Y/N	N	FAIR
Schatcher, 94	Y	Y/N	Y	GOOD

Appendix F: Included Studies by Individual Factor

Categorical Data Factors

Factor	Reference Numbers
Past Obstetric	
Parity of one	17, 83, 87
Number of prior CD	
One prior CD (versus more than one)	9, 10, 47, 50, 51, 55, 65, 92, 100, 102, 106, 109, 125, 131
One prior CD (versus two prior CD)	10, 65, 100, 106, 109, 125
One prior CD (versus three prior CD)	10, 65, 100, 106, 109, 125
Two prior CD (versus three prior CD)	10, 65, 81, 100, 106, 109, 113. 125
Previous VD	24, 50, 52, 53, 55, 57, 58, 61, 66, 81, 90-92, 94, 98, 109, 111, 118, 120
Previous VD order	
Before prior CD (versus after)	24, 59, 68, 84, 87, 94, 110
Before prior CD (versus none before)	24, 68, 84, 87, 94, 110
After prior CD (versus none after)	24, 68, 84, 87, 92, 94, 110
Indication of prior CD	
Recurrent (versus non-recurrent)	9, 10, 26, 47, 48, 50, 51, 57, 60-63, 65, 66, 69, 70, 74, 75, 79-82, 84, 85, 87, 90, 94, 95, 97, 99, 102, 106, 107, 110-112, 117-120, 123, 125-127, 129-131, 133, 134, 139
Breech (versus recurrent)	9, 10, 26, 48, 50, 51, 57, 60, 63, 70, 75, 75, 81, 82, 84, 85, 87, 90, 94, 97, 99, 102, 107, 110-112, 118, 119, 126, 129, 130, 133, 134, 139
Fetal Distress (versus recurrent)	9, 10, 26, 48, 51, 57, 60, 63, 70, 75, 75, 81, 82, 84, 85, 90, 94, 97, 99, 102, 107, 110-112, 118, 119, 126, 129, 130, 133, 134, 139
Previous Cervical Dilation <4cm	87, 90, 127
Current Obstetric	
Spontaneous labor	55, 56, 59, 61, 62, 64, 65, 84, 90, 92, 94, 98, 105, 114, 116, 121, 123, 137, 140
Oxytocin use (non-specified)	9, 10, 62, 75, 83, 84, 102, 105, 107, 110, 111, 116, 117, 125, 126, 131, 134, 135
Augmented labor	55, 58, 62, 65, 69, 79, 84, 90, 92, 94, 113, 114, 116, 123, 127, 130, 140
Epidural use	48, 55, 72, 92, 94, 98, 102, 105, 107, 117, 125, 140
Birth weight <4000g	17, 62, 73, 87, 94, 102, 105, 107, 114, 118, 123, 138

Appendix F: Included Studies by Individual Factor – Continued

Continuous Data Factors

Factor	Reference Numbers
Demographic	
Age (yrs)	9, 23, 26, 49, 61, 92, 94, 102, 108, 116, 122, 135, 140
Past Obstetric	
Gravidity	23, 26, 92, 116, 117
Parity	26, 102, 116, 117, 122
Current Obstetric	
Gestational age (wks)	23, 49, 61, 92, 94, 102, 116, 117, 122, 135
Admission cervical dilation (cm)	92, 116, 117
Birth weight (g)	26, 60, 61, 64, 49, 68, 88, 91, 108, 116, 117, 122, 129, 134
Maternal height (cm)	49, 91, 94, 108, 135
Maternal weight (kg)	94, 102, 135

Appendix G: Random Effects Modeling for Categorical Variables using WINBUGS®

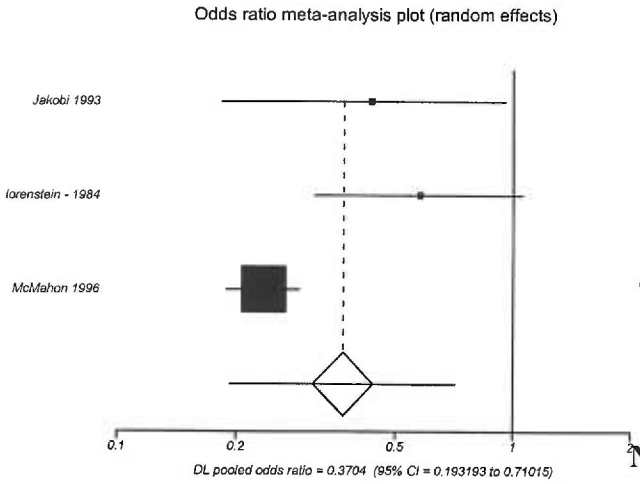
Factor	Number of Studies	Average VBAC proportion*	Odds Ratio for VBAC	95% CI
Past Obstetric				
Parity of one	3	55.9	**0.267**	**0.223,0.319**
Number of prior CD				
One prior CD (versus more than one)	14	80.2	**1.560**	**1.414,1.725**
One prior CD (versus two prior CD)	6	82.8	**1.628**	**1.453,1.705**
One prior CD (versus three prior CD)	6	82.8	1.205	0.874,1.705
Two prior CD (versus three prior CD)	8	75.1	0.775	0.525,1.046
Previous VD	20	88.8	**2.92**	**2.570,3.296**
Previous VD order				
Before prior CD (versus after)	7	80.5	**0.375**	**0.297,0.477**
Before prior CD (versus none before)	6	82.6	**1.481**	**1.254,1.743**
After prior CD (versus none after)	7	93.5	**4.655**	**3.638,5.880**
Indication of prior CD				
Recurrent (versus non-recurrent)	49	64.5	**0.531**	**0.504,0.558**
Breech (versus recurrent)	34	84.1	**3.012**	**2.775,3.253**
Fetal Distress (versus recurrent)	32	72.8	**1.528**	**1.412,1.653**
Previous Cervical Dilation <4cm	3	68.9	0.844	0.506,1.320
Current Obstetric				
Spontaneous labor	19	76.2	**1.672**	**1.557,1.792**
Oxytocin use (non-specified)	18	67.6	**0.467**	**0.428,0.510**
Augmented labor	16	79.5	**1.423**	**1.330,1.525**
Epidural use	12	75.8	**0.814**	**0.677,0.965**
Birth weight <4000g	12	73.3	**1.665**	**1.541,1.798**

Bold = significant difference; * average VBAC proportion in those with that factor

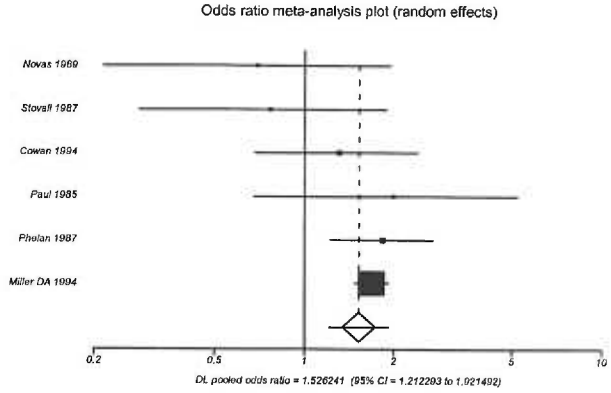
Appendix H: Forest Plots – by sample size (smallest to largest)

* significant heterogeneity ($p<0.10$) by Breslow-Day test statistic

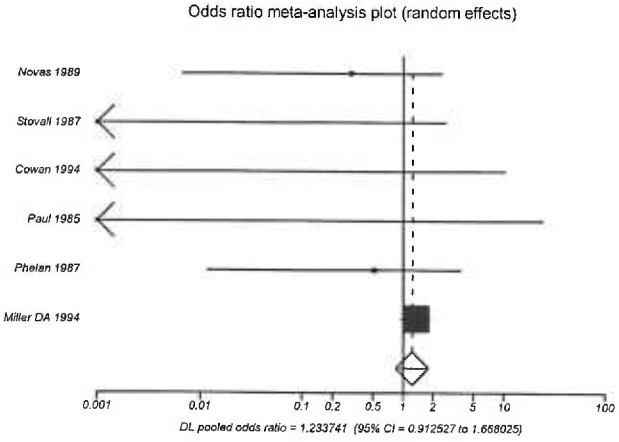
Parity: one versus more than one*



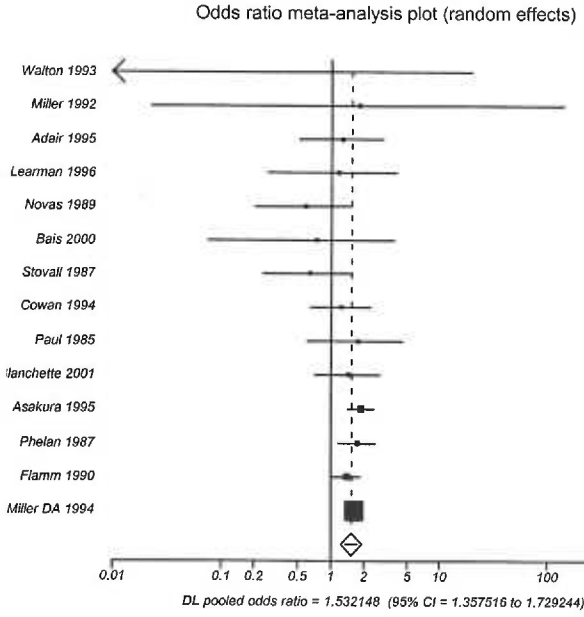
Number of prior CD: one versus two



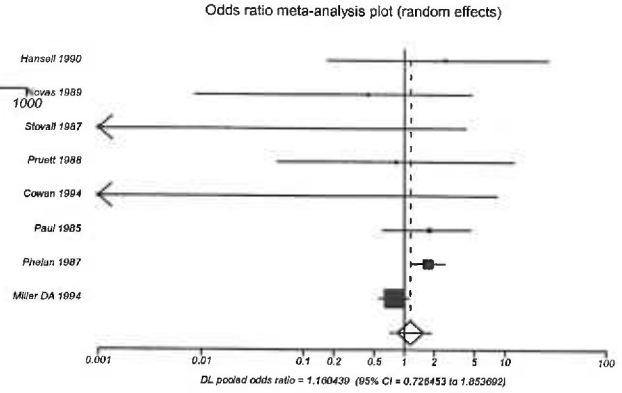
Number of prior CD: one versus three



Number of prior CD: one versus more than one



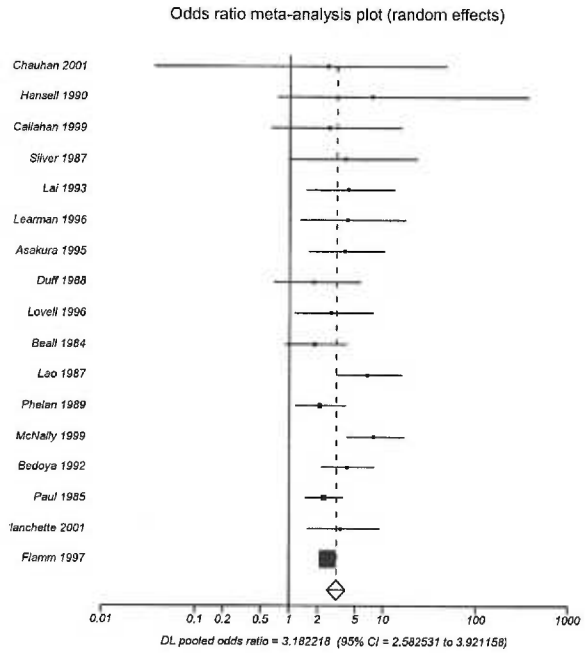
Number of prior CD: two versus three*



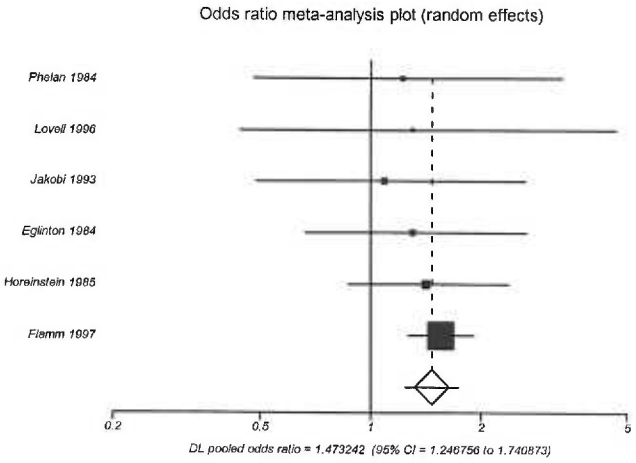
Appendix H: Forest Plots – Continued

* significant heterogeneity ($p<0.10$) by Breslow-Day test statistic

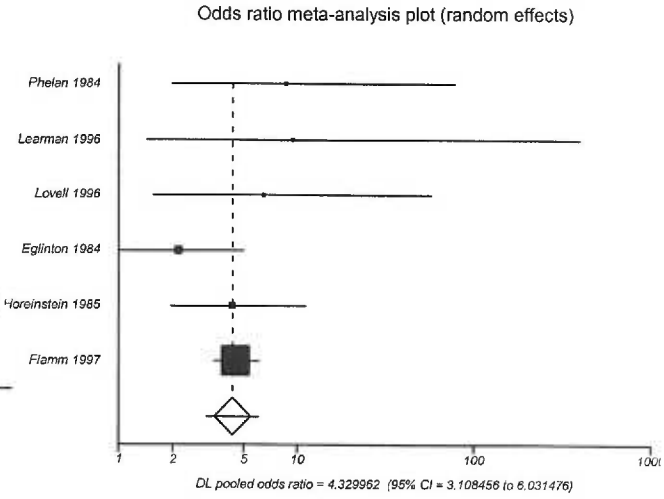
Previous VD*



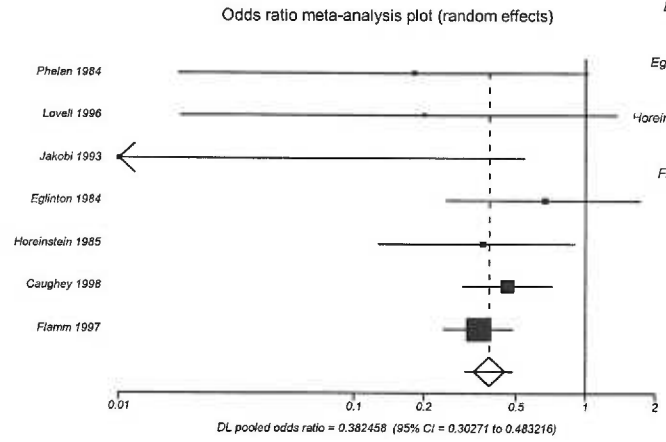
PVD order: before versus none before



PVD order: after versus none after



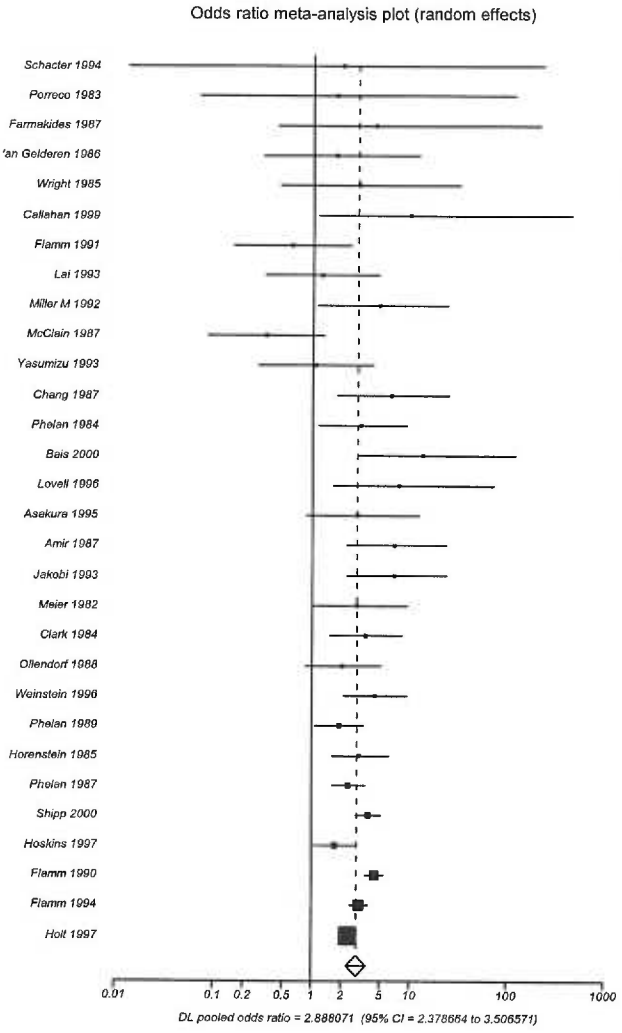
PVD order: before versus after



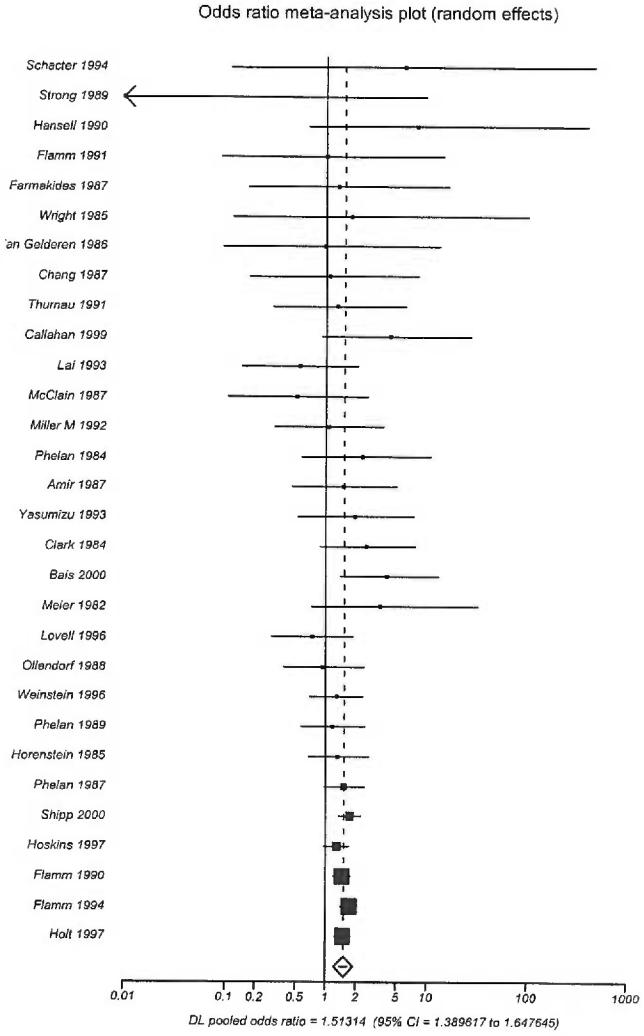
Appendix H: Forest Plots – Continued

* significant heterogeneity ($p<0.10$) by Breslow-Day test statistic

Prior CD indication: breech*



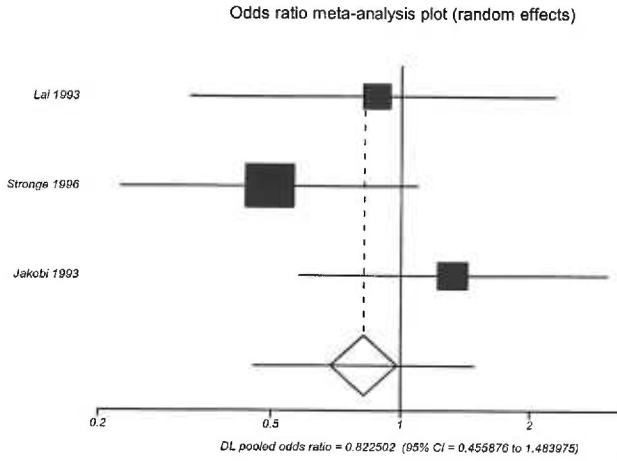
Prior CD indication: fetal distress



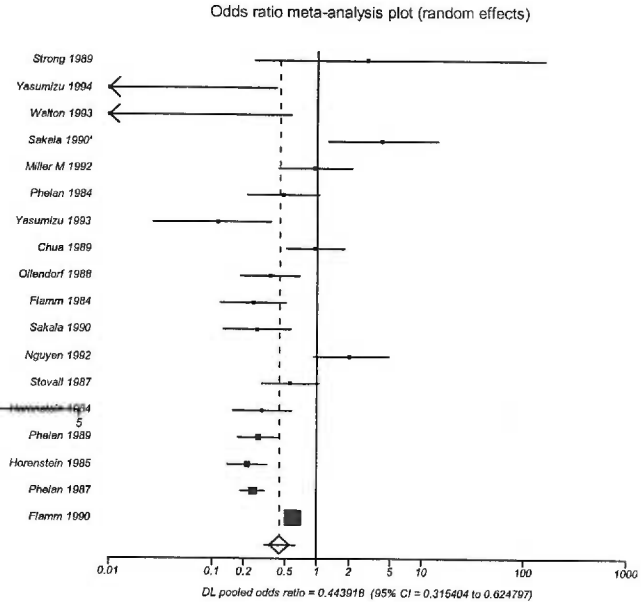
Appendix H: Forest Plots – Continued

* significant heterogeneity ($p<0.10$) by Breslow-Day test statistic

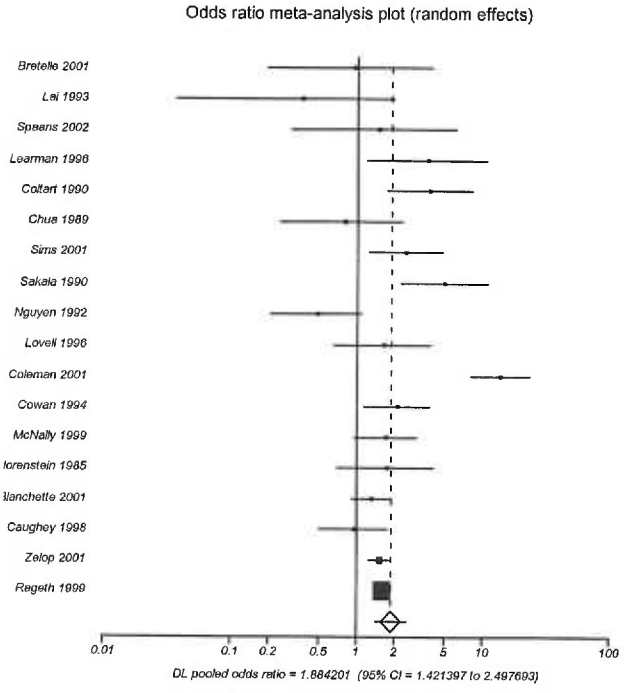
Previous Cervical Dilation <4cm



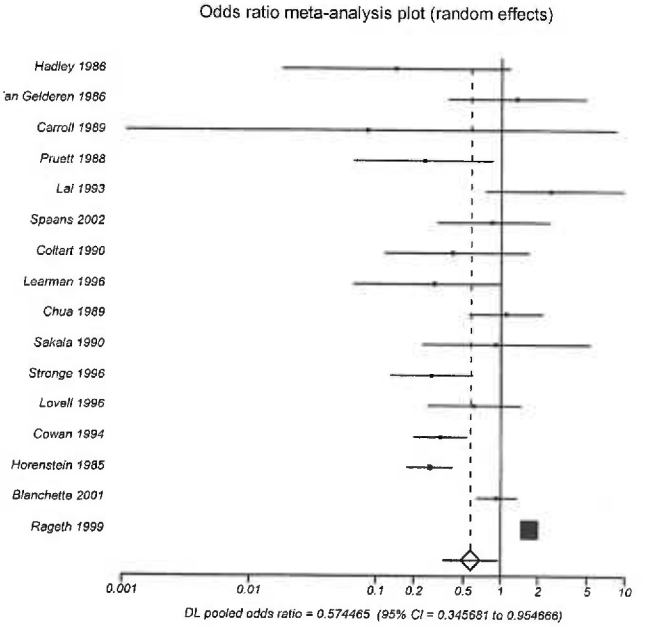
Oxytocin use (non-specified)*



Spontaneous Labor*



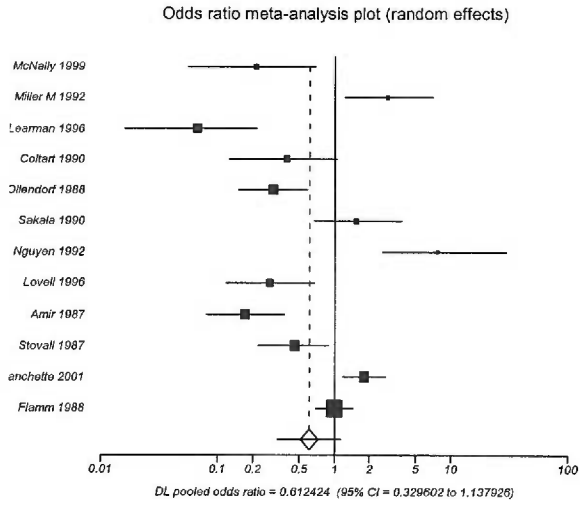
Augmentation*



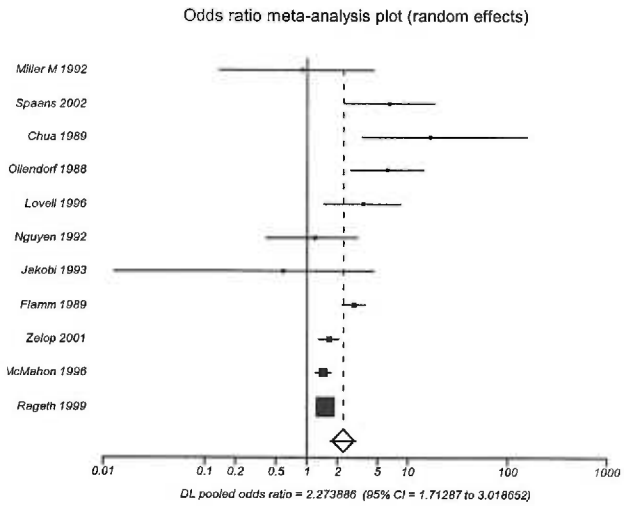
Appendix H: Forest Plots – Continued

* significant heterogeneity (p<0.10) by Breslow-Day test statistic

Epidural use*



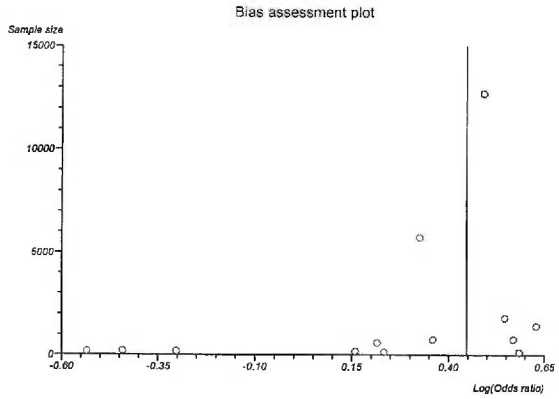
Birthweight <4000 grams*



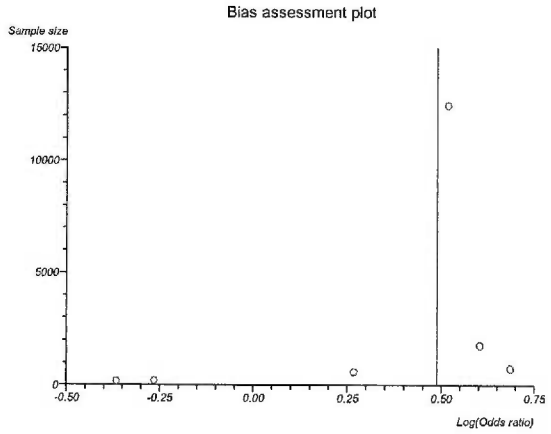
Appendix I: Funnel plots for the Detection of Publication Bias

Parity: one versus more than one – insufficient number of studies

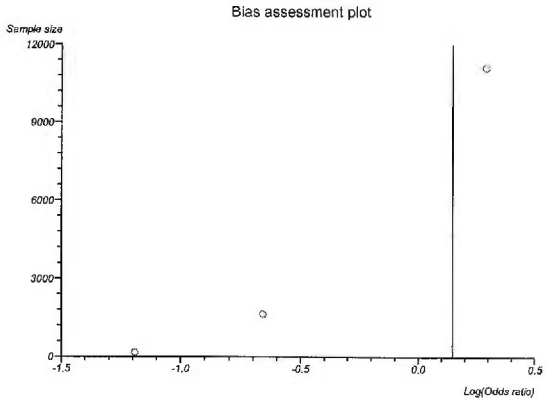
Number of prior CD: one versus more than one



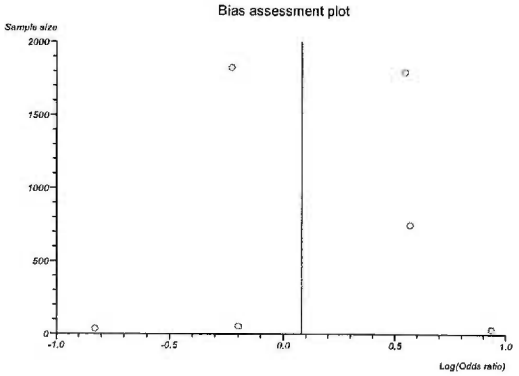
Number of prior CD: one versus two



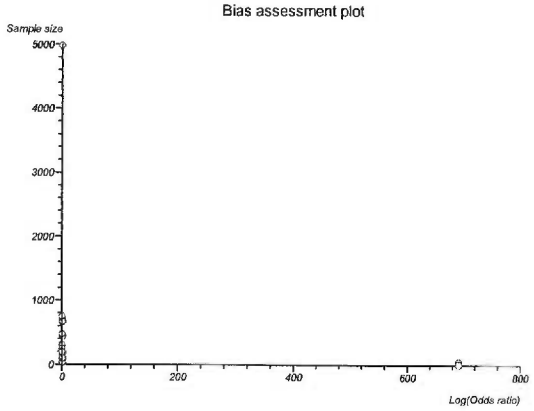
Number of prior CD: one versus three



Number of prior CD: two versus three

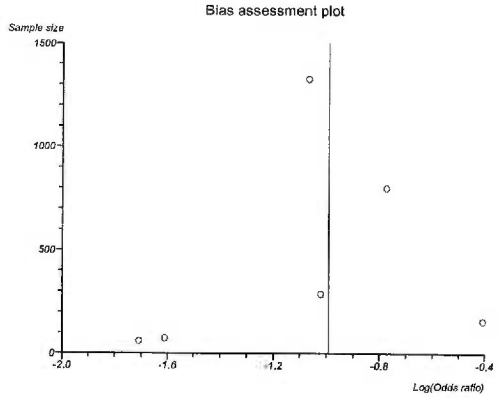


Previous VD

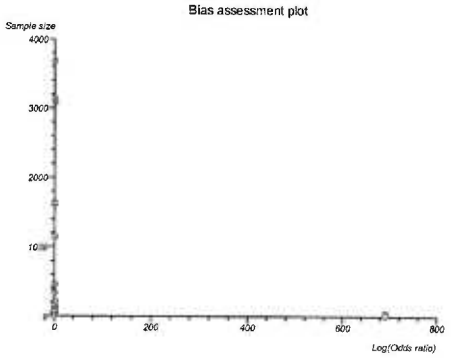


Appendix I: Funnel plots for the Detection of Publication Bias - Continued

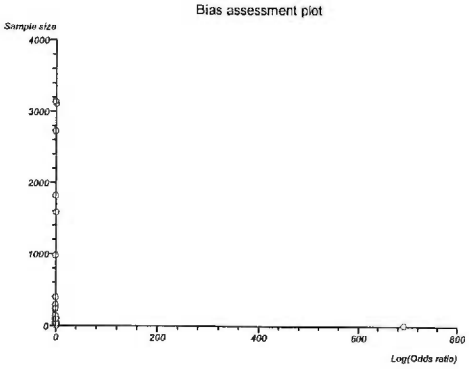
PVD order: before versus after



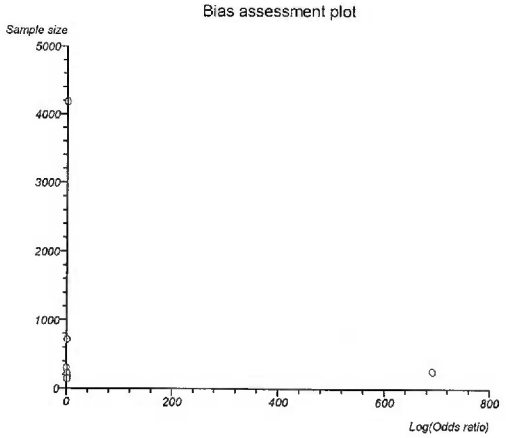
Prior CD indication: breech



Prior CD indication: fetal distress



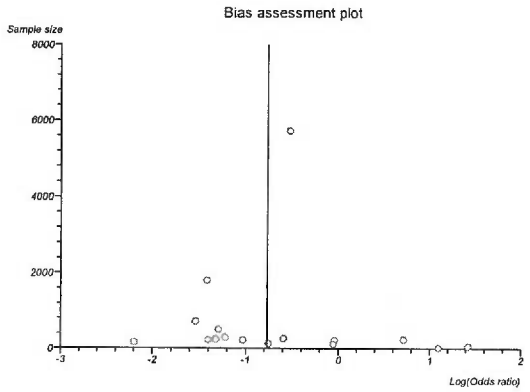
PVD order: after versus none after



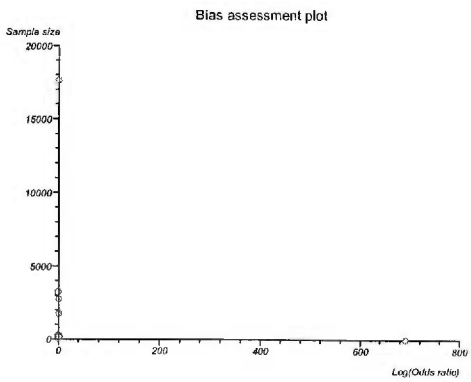
Previous Cervical Dilation <4cm - insufficient number of studies

Appendix I: Funnel plots for the Detection of Publication Bias - Continued

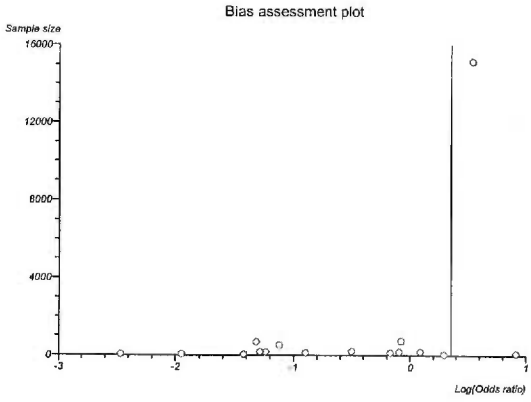
Oxytocin use (non-specified)



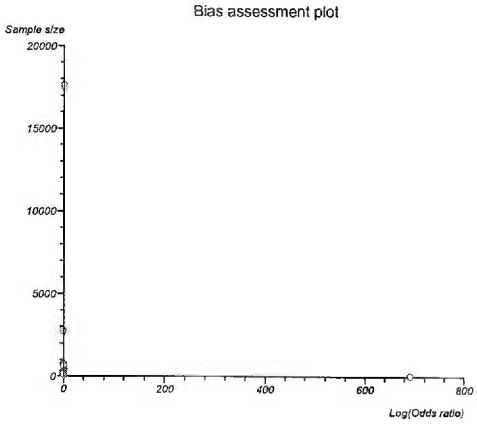
Birthweight <4000 grams



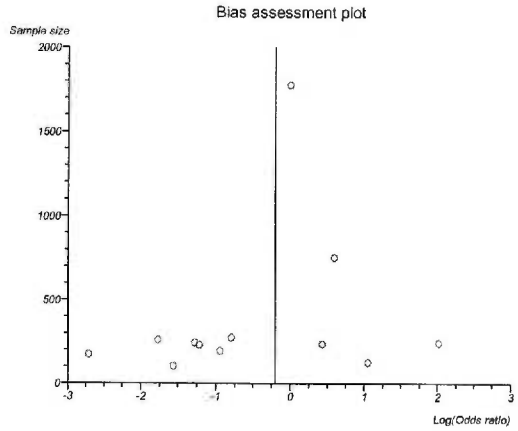
Augmentation



Spontaneous Labor



Epidural use



Appendix J: MOOSE Guidelines

1. Reporting of background should include
    a. Problem definition
    b. Hypothesis statement
    c. Description of study outcome(s)
    d. Type of exposure or intervention used
    e. Type of study designs used
    f. Study population
2. Reporting of search strategy
    a. Qualifications of searchers (e.g. librarian and investigators)
    b. Search strategy, including time period included in the synthesis and keywords
    c. Effort to include all available studies, including contact with authors
    d. Databases and registries searched
    e. Search software used, name and version, including special features used (e.g. explosion)
    f. Use of hand searching (e.g. reference lists of obtained articles)
    g. List of citations located and those excluded, including justification
    h. Method of addressing articles published in languages other than English
    i. Method of handling abstracts and unpublished studies
    j. Description of any contact with authors
3. Reporting of methods should include
    a. Description of relevance or appropriateness of studies assembled for assessing the hypothesis to be tested
    b. Rationale for the selection and coding of data (e.g. sound clinical principles or convenience)
    c. Documentation of how data were classified and coded (e.g. multiple raters, blinding, and interrater reliability)
    d. Assessment of confounding (e.g. comparability of cases and controls in studies where appropriate)
    e. Assessment of study quality, including blinding all quality assessors; stratification or regression on possible predictors of study results
    f. Assessment of heterogeneity
    g. Description of statistical methods (e.g. complete description of fixed or random effects models, justification of whether the chosen models account for predictors of study results, dose-response models, or cumulative meta-analysis) in sufficient detail to be replicated
    h. Provision of appropriate tables and graphs
4. Reporting of results
    a. Graphic summarizing individual study estimates and overall estimate
    b. Table giving descriptive information for each study included
    c. Results of sensitivity testing (e.g. subgroup analysis)
    d. Indication of statistical uncertainty in findings
5. Reporting of discussion should include
    a. Quantitative assessment of bias (e.g. publication bias)
    b. Justification for exclusion (e.g. exclusion of non-English language citations)
    c. Assessment of quality of included studies
6. Reporting of conclusions should include
    a. Consideration of alternative explanation of results
    b. Generalization of the conclusions (e.g. appropriate for the data presented and within the domain of the literature review)
    c. Guidelines for future research
    d. Disclosure of findings

Appendix K: Abbreviations and Definitions

Abbreviations:

a. BW – birthweight
b. CD – cesarean delivery
c. CPD – Cephalopelvic disproportion
d. ECV – external cephalic version
e. EFW – estimated fetal weight
f. ERCD – elective repeat cesarean delivery
g. FHT – fetal heart tracing
h. FTOL – failed trial of labor
i. FTP – failure to progress
j. GA – gestational age
k. IUGR – intrauterine growth restriction
l. LTCS – lower segment transverse cesarean section
m. NA – not applicable
n. NPV – negative predictive value
o. NS – NR – non significant / actual p-value not reported
p. OR (a) – adjusted odds ratio
q. Prior CD – previous cesarean delivery
r. PPV – positive predictive value
s. RCT – randomized clinical trial
t. SES – socioeconomic status
u. TOL – trial of labor
v. VBAC – vaginal birth after cesarean
w. VD – vaginal delivery

Definitions:

a. Recurrent prior CD indication: cephalopelvic disproportion (CPD), failure to progress (FTP)
b. Non-Recurrent prior CD indication: all other indications

OREGON HEALTH SCIENCES UNIVERSITY

SCHOOL OF MEDICINE GRADUATE STUDIES PROGRAM

THESIS PLACEMENT WAIVER

Student Name: Jason Naoki Hashima

I grant permission to the Oregon Health Sciences University to place a copy of my thesis entitled,

Predicting Vaginal Birth After Cesarean Success: Is it Possible?

in the main University library and in the departmental or interdisciplinary graduate program collection of theses.

Student — 3/28/03 date

Associate Dean for Graduate Studies — date

revised 9/98