ASSESSING THE SAFETY OF FREESTANDING BIRTH CENTERS WITH PROPENSITY SCORE ANALYSIS

By

Heather A. Davidson

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

December, 2005

Nashville, Tennessee

Approved:

Professor David S. Cordray

Professor Bruce Compas

Professor Georgine Pion

Professor Melanie Lutenbacher

Professor Colleen Conway-Welch

ACKNOWLEDGEMENTS

This project could not have been possible without the financial support of the W.K. Kellogg Foundation and the Kellogg Birth Center Evaluation Project. I sincerely thank Dr. David S. Cordray for giving me the opportunity to work with and learn from him on such an extensive evaluation. Dr. Cordray's guidance through the entire dissertation process has been tremendous, and my professional growth from this experience is immeasurable. I also thank my Dissertation Committee members, Dr. Bruce Compas, Dr. Georgine Pion, Dr. Melanie Lutenbacher, and Dr. Colleen Conway-Welch, for their excellent observations and professional advice. And, I thank Dr. Warren Lambert for his consultations and assistance navigating propensity score analysis.

I also offer many thanks to the nurse-midwives and staff at each freestanding birth center in Tennessee who allowed me to learn from their experience. Most especially, I express tremendous appreciation to Margaret Buxton, Linda Cole, Cliff Honicker, and Liz Howard for their integral role in my education. They have enriched my understanding of what it means to provide excellent care during pregnancy and what it takes to work successfully in the maternal health care system.

And finally, I have everlasting gratitude for the steadfast support of my dear family and many friends throughout this chapter of my life. To my parents, Dr. Glen and Shirlee Davidson, I thank you for your infinite confidence in my abilities; and to my sister, Kristin Davidson, for always reminding me to fly. And to Michael Stahl, I thank you for your endless love and support that keeps me laughing, focused, and charging forward.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

A primary goal for all prenatal and childbirth services is to maximize the health of women and infants, and therefore, provide safe care. Obstetrical texts define safe care as the ability to identify risks and appropriately intervene in order to reduce impact of the risks on the health of mothers and their infants (Bernstein, Harrison, & Merkatz, 2000). Despite this common goal, mainstream professional opinion in the United States suggests that no out-of-hospital facility for childbirth can be safe because appropriate intervention is not available (Cunningham et al., 1997). Freestanding birth centers (FBCs), out-of-hospital services that provide prenatal and childbirth care for low medical risk women, are therefore regularly scrutinized by professionals to provide evidence that their services are as safe as those obtained in hospital settings. The purpose of this study is to evaluate the relative safety of obtaining prenatal and childbirth services at FBCs by utilizing methodological strategies to control for hidden biases common in nonrandomized, observational designs.

The freestanding birth center has functioned as an alternative childbearing site to both home and standard hospital care for over 30 years. According to the National Association of Childbearing Centers (NACC):

> A birth center is a homelike facility, existing within a healthcare system with a program of care designed in the wellness model of pregnancy and birth. Birth centers are guided by principles of prevention, sensitivity, safety, appropriate medical intervention, and cost effectiveness. Birth centers provide family-centered care for healthy women before, during and after normal pregnancy, labor and birth (NACC, 1999).

Nurse-midwives are the primary providers at FBCs and their philosophy of care focuses on the normalcy rather than the pathology of childbirth (Kennedy, 2000, 2002; Varney, Kriebs, & Gegor, 2004). A principle of the midwifery philosophy of care is also the belief that a collaborative relationship must exist between the woman and her provider during the birth experience, as compared to decision-making that is controlled mostly by the provider (Kennedy, 2000; Raisler, 2000).

Risk assessment occurs continuously throughout a woman's pregnancy at NACC accredited freestanding birth centers. If problematic risks occur during the pregnancy, providers at the FBCs transfer the woman to hospital-based care. If a woman remains at a low medical risk throughout her pregnancy, she is admitted to the FBC for childbirth. In the event of a complication during delivery, accredited FBCs are located within a 30-minute transfer radius to local hospitals. Furthermore, all centers are required to have medical directors to help monitor more complex cases (NACC, 1999). For these reasons, evaluations of safety at FBCs tests care both the prenatal referral process and childbirth care (Rooks, 1997).

Evaluating whether FBCs are a safe option for care requires investigating whether providers of this service can effectively minimize prenatal and childbirth risks compared to services obtained through a hospital system. Ideally, a randomized controlled trial would answer this research question by controlling for all observed and unobserved risks associated with participants in both services. However, because of ethical considerations pertaining to an individual's needs and choices, randomized trials are rarely implemented in childbirth services research. The majority of FBC studies have relied on nonrandomized matching techniques and statistical adjustments in observational studies

to evaluate outcomes from comparable groups of women choosing FBCs or hospital services.

Comparative studies have consistently found non-significant differences in the majority of important infant and maternal health outcomes between FBCs and hospital services (Jackson et al., 2003; Walsh & Downe, 2004). However, the validity of this evidence is often questioned because of methodological limitations, such as inadequately designed comparison groups that are susceptible to selection bias and small sample sizes that are not powerful enough to detect significant differences of rarely occurring outcomes (Albers & Katz, 1991; Fullerton & Wingard, 1990; Jackson, Lang, Dickinson, & Fullerton, 1994; Lieberman & Ryan, 1989; Walsh & Downe, 2004). When sample sizes are too small, statistical non-significance does not necessarily imply statistical equivalence.

The major challenge to evaluating the safety of childbirth services in a nonrandomized design is that study groups must have comparable levels of risk. Designing hospital comparative groups is difficult because FBCs care for overall healthier women than women with low risk pregnancies using hospital services (Institute of Medicine & National Research Council [IOM/NRC], 1982; Albers & Katz, 1991). To address this concern and add to the current understanding of FBC safety requires designing a comparative study that takes into account the complexities associated with achieving comparable levels of risk. These complexities include taking into account multiple factors within the biopsychosocial model of risk, where the underlying mechanisms causing poor childbirth outcomes, such as low birth weight and preterm birth, are made up of multiple paths that can produce the same outcome (Paneth, 1995).

Furthermore, in order to design appropriate in-hospital comparison groups for a FBC, underlying risks associated with psychosocial factors must also be considered. Past research has found that women have distinct attitudes and beliefs driving their choice of an out-of-hospital birth (Howell-White, 1999). Therefore, psychosocial differences may factor into how women cope with their pregnancies and have the potential to drive health behaviors. Lastly, known differences in provider practices may influence outcomes. It is suggested in empirical literature that provider practices between nurse-midwives and obstetricians are different with respect to the number of interventions (Rosenblatt et al., 1997). The number of interventions used during pregnancy influence outcomes either for better, such as avoiding problems, or worse, leading to risks associated with unnecessary procedures. All of these factors have the potential to introduce bias into a non-random comparison study between childbirth services, and therefore, are considered in the design of this study.

The current study evaluated the safety of FBCs by employing several methodological strategies specifically designed to minimize the bias attributed to differences in overall risk profiles, patient choice, and provider practice between FBC and hospital services. Using methodological frameworks presented by Rosenbaum (2002) and Rubin (1997), the goal of this study was to create comparison groups that mimic experimental control found in randomized comparison groups, thus, creating a counterfactual condition from which to compare the outcomes for FBC participants. Creating a counterfactual condition minimized the influence of biases and helps estimate whether there is a causal effect of service associated with pregnancy outcomes. In particular, the current study: (1) selected only those women who obtain provider services

from nurse-midwives in hospitals geographically similar to those in FBCs; and (2) used propensity score analysis to balance multiple risks observed on the birth certificates (Rubin, 1997).

Results from this study contribute to both research and practice in the field of childbirth services. From a research perspective, the opportunity to estimate safety by using methodologies that minimize the impact of hidden biases help inform the debate over how to better evaluate these services. From a practice perspective, the viability of FBC services as part of the larger maternal care system hinges on providing evidence of a safe childbirth experience (IOM/NRC, 1982). Although less than 0.5% of all births in the United States occurred at FBCs in 2002 (Martin et al., 2003), the availability of alternative services for pregnant women plays an important role, given the decreasing availability of obstetrical services in rural areas. Services such as FBCs are one option for meeting the needs of vulnerable populations, and evaluating their safety affects women's health on both individual and policy levels.

CHAPTER II

LITERATURE REVIEW

Historical Context

The appropriate location for childbirth has been a topic of debate for researchers, clinicians, and social advocates for at least 150 years (Davis-Floyd, 2001; Devitt, 1996; Howell-White, 1999; Leavitt, 1999; Rooks, 1997); a debate which continues to exist in the experience of individual families (Davidson, Cordray, & Johnson, 2002). Between 1930 and 1960, the number of women giving birth in hospitals rose dramatically from 37% to 96% nationally (Devitt, 1996), and the majority of births occur in hospitals today (Martin et al, 2003). Many have credited this population-wide move into the hospital as the primary cause for improvement in childbirth outcomes (Cunningham et al., 1997). However, it is very difficult to unravel the effect of hospital services on morbidity and mortality rates of pregnancy when advances in medical training, availability of antibiotics and other pharmaceutical interventions, nutrition standards and improved living conditions have also contributed to the outcomes (Devitt, 1996; Maine, 1991). In fact, obstetrical specialists were presented the "wooden spoon", a dubious distinction recognizing missed opportunities in science, by Archie Cochrane of the Cochrane Library because they had missed opportunities to conduct randomized trials on home versus hospital birth sites for low risk women in the 1960's (Olsen & Jewell, 2000).

Although hospital births were becoming the norm for Americans, messages from the deinstitutionalization and feminist movements of the 1960's and 1970's encouraged

consumers to voice their discontent with the norms of the hospital system, and people advocated for more "humane, less costly, and less centralized methods of health care" (Eakins, 1984). Discontent with the hospitalized birth process continues as maternal health professionals criticize hospital-based obstetricians for treating all pregnancies as high-risk (Wagner, 2001). Critics suggest that by making "high-risk" the norm, hospital protocols contribute to an increasing probability of unnecessary interventions such as induction and caesarian section, technologies without proper evidence of their effectiveness, and lack of psychosocial support available to women during childbirth (Rooks, 1997; Leavitt, 1999). These messages have been part of the societal backdrop during development of the current freestanding birth center concept.

One of the first documented birth centers in the United States, La Casita, was developed in the 1940's by nurse-midwives associated with the Catholic Maternity Institute of Santa Fe, New Mexico to serve the rural poor in need of home-like facilities closer to a hospital (Rooks, 1997). The majority of FBCs, however, rose out of the home birth movement of the 1960's, when opponents of hospital births felt that the prevailing system had failed to "meet the needs of childbearing families for personalized, health oriented services" (Lubic & Ernst, 1978). The Maternity Center Association in New York City opened the first urban FBC, The Childbearing Center, in 1975, as a demonstration project designed to meet the needs of families opposed to giving birth in hospitals. Lubic and Ernst (1978) concluded that The Childbearing Center in New York provided a safe, satisfying, and economical out-of-hospital birth experience. Goals of The Childbearing Center aligned closely with the goals of the American Public Health Association which in 1979, supported the "development of demonstration projects for

alternatives in maternity care," and research on "family centered maternity care" (American Public Health Association, 1983).

The rising number of out-of-hospital births in the 1970's prompted the Institute of Medicine's Office for Maternal and Child Health (OMCH) to form the Committee on Assessing Alternative Birth Settings. The Committee evaluated the state of the research literature, finding mostly descriptive studies that reported low rates for poor outcomes among out-of-hospital populations. They concluded, however, that lack of randomized control designs and measurement limitations impeded any statistically conclusive decision regarding the safety of alternative birth sites. Methodological limitations included lack of clearly identified risks, lack of quantified risks, no adequate comparison studies, inadequate small sample sizes, and studies that poorly controlled for confounding factors (Institute of Medicine & National Research Council, 1982). Other reviews, consistent with the IOM/NRC conclusions, suggested that little objective data provided clear advantages or disadvantages for any birth site, hospital or home (Adamson & Gare, 1980). The IOM (1982) report strongly suggested that research needed to identify the differences, if any, between alternative settings and more conventional hospital birth sites.

By the late 1980's, the American College of Obstetricians and Gynecologists (ACOG), and the American Academy of Pediatrics (AAP) released statements criticizing the lack of research assessing safety of freestanding birth centers. These national organizations recommended the hospital as the safest childbearing site (AAP/ACOG, 1988). Because of the methodological critiques in the research literature, the safety of FBCs remains an open debate. Present day critics of FBCs are concerned that any out-of-

hospital childbirth birth site will not have the adequate technology or trained professionals to safely intervene with unexpected problems that may lead to avoidable morbidity and mortality (Cunningham et al., 1997; Mathews & Zadak, 1991).

Research on the Safety of Freestanding Birth Centers

Researchers conducting studies in response to these recommendations have consistently found low levels of poor pregnancy outcomes at FBCs. The National Birth Center Study (Rooks, Weatherby, & Ernst, 1992a, 1992b, 1992c; Rooks et al., 1989) reported outcomes for 11,814 births at 84 FBCs in the United States. This large study collected risk and outcome data at three different time-periods: First prenatal visit, labor and delivery, and postpartum care. In their sample, the infant mortality rate was 1.3 out of 1,000 births; there was no instance of maternal mortality; 2.4% of women experienced preterm birth; 0.8% of babies were born with low birth weight; and the cesarean section rate was approximately half the reported rate as that for low-risk, in-hospital births in previous studies of national samples. Although much less extensive and using smaller sample sizes, several other descriptive studies in the same time period reported similar low rates of poor childbirth outcomes for women using FBCs (Bennetts & Lubic, 1982; DeJong, Shy, & Carr, 1981; Eakins, O'Reilly, May, & Hopkins, 1989; Faison, Pisani, Douglas, Cranch, & Lubic, 1979; McCallum, 1979; Reinke, 1982). Few of the studies during the early to mid-1980's, however, utilized comparison group designs; therefore, they were unable to meet the methodological requirements set forth by the IOM report (1982) that support statistical conclusiveness on safety.

Two recent reviews using meta-analytic procedures looked at studies with comparative group designs to statistically compare childbirth outcomes between FBC and hospital services (Davidson, 2001; Walsh & Downe, 2004). Davidson (2001) found five eligible comparison studies that reported outcomes relevant to issues of safety (Baruffi et al., 1984; Feldman & Hurst, 1987; Fullerton & Severino, 1992; Scupholme, McLeod, & Robertson, 1986; Waldenstrom, Nilsson, & Winbladh, 1997). However, synthesizing outcomes was difficult because the reviewed studies did not operationalize safety using the same variables. Low birth weight and an Apgar score of less than seven at five minutes were the only outcomes measured consistently across all studies. Other measurable outcomes included incidences of shoulder dystocia, hemorrhages, meconium present in the amniotic fluid, rates of maternal fever, gestation length less than 37 weeks, prolonged gestation (42+ weeks gestation), infant resuscitation, neonatal mortality rate, neonatal morbidity scales, and the absence of overall complications. Table 1 summarizes the odds ratios and mean effect sizes calculated for this review.

The majority of effect sizes calculated for each outcome were non-significant between FBC and hospital groups. Outcomes with significant effect sizes, such as maternal fever (OR: 0.37, CI 0.22-0.64) and a neonatal morbidity score (OR: 0.63, CI 0.49-0.81), both had higher incidences in the standard hospital group. For the purpose of the synthesis, an aggregated effects size reported for major morbidity and mortality indicators was also calculated resulting in a non-significant difference (OR: 0.963) between FBC and hospital groups. Although all of the studies used matched comparison designs, none of them adjusted for differing risk profiles or demographics.

Table 1: Effect Sizes Calculated from Davidson (2001), Comparing Outcomes at Hospitals to FBCs

Maternal and Infant complications during childbirth	Aggregated effect sizes (Odds Ratios with 95% Confidence Intervals)
Meconium present in amniotic fluid (E)	0.70 (0.47-1.04)
Maternal fever (D)	0.37 (0.22-0.64)*
Shoulder dystocia (E, H)	0.47 (0.15-1.50)
Hemorrhage (E, H)	0.62 (0.76-4.51)
Complications as a function of 42+ weeks (H)	0.96 (0.84-1.09)
Gestation length < 37 weeks (J)	1.35 (0.81-2.23)
Birth weight < 2500 g (D, E, F, H)	0.68 (0.44-1.06)
Infant resuscitation (D)	1.42 (0.88-2.28)
Neonatal mortality rate (D)	0.42 (0.11-1.61)
Neonatal morbidity score > 10 (D)	0.63 (0.49-0.81)*
No complications present (J)	0.92 (0.74-1.15)
Apgar at 1 minute < 7 (D, F)	0.014 (-0.012-0.039) mean effect size
Apgar at 5 minutes < 7 (D, E, H, J)	0.024 (0.021-0.026) mean effect size

D: Baruffi et al. (1984); E: Scupholme et al. (1986); F: Feldman, E. & Hurst, M. (1987); H: Fullerton & Severino (1992); J: Waldenstrom et al. (1997)
*$p<0.05$

More recently, a large comparison study added to research assessing the safety of a collaborative nurse midwife and obstetrician management group with a San Diego-based FBC or hospital option for delivery (Jackson et al., 2003). Jackson et al. sampled low-income pregnant women who chose to give birth at a FBC (n = 2756) with collaborative care or at a hospital (n = 1577) with traditional care. To address the critique of self-selection bias, they used birth center eligibility criteria to prospectively match traditional care participants who would have been eligible for FBC services based on medical risk. Their sample size had the statistical power to detect a 3-5% difference for major outcomes between groups. The Jackson et al. study statistically adjusted for other confounding risks, including race/ethnicity, parity, c-section history, education, age, marital status, country of origin, height, and smoking during pregnancy. They found that, even though groups differed on demographic factors, statistical adjustment on these factors did not alter the results.

Results from the Jackson et al. (2003) study indicate that there were non-significant differences in proportions for major outcomes between FBC and hospital deliveries including major neonatal complications (adjusted risk difference: -1.8, 95%CI: -3.8, 0.1), Apgar score under 7 at 5 minutes (0.9, 95% CI: -3.7,5.4), preterm delivery (0.2, 95% CI: -1.7,2.1), low birthweight (0.5, 95% CI: -1.7, 2.7), neonatal intensive care unit admission (-1.3, 95% CI: -3.8,1.1), or neonatal readmission after 28 days of age (-1.3, 95% CI: -4.1, 1.5). Similarly, there were no significant differences in the adjusted proportions of major maternal antepartum complications (-0.5, 95%CI: -2.5, 1.5), intrapartum complications (0.8, 95% CI: -2.4, 4.0), or postpartum complications (0.6, 95% CI: -4.2, 5.3). The only significant difference in proportion between services was the number of infants needing sepsis assessment after 1-3 days, which was a higher risk for the hospital group (relative risk ratio = -3.8; CI: -6.4-1.3).

The Kellogg Birth Center Evaluation project (Cordray & Davidson, 2004) is the most recent study assessing safety of FBCs. The evaluation was part of a W.K. Kellogg grant awarded to the Vanderbilt University School of Nursing to develop a network of FBCs in Tennessee between 1999 and 2003. The evaluation assessed the efficacy of three FBCs, two urban-based services and one rural-based service. The goals of this project were to evaluate the use of FBC services as a possible method to increase access to quality prenatal care, and provide a safe birthing alternative to women with low risk pregnancies in medically underserved areas.

The evaluation assessed the FBCs for safety, using birth certificate records and abstracted information from medical charts. The study developed a sampling frame for hospital comparison groups, in similar geographical areas for each FBC, by locating birth

certificate records where hospital births were midwife-attended. Using midwife-attended hospital births for the comparison group specifically minimized the potential influence of selection factors associated with choosing midwifery philosophy of care and potential biopsychosocial and socio-economic differences. In order to reduce the concern that selection bias influences the accuracy of results, the objective was to create comparison groups of women who differed as little as possible from the FBC groups. The current study built off of this design by also using midwife-attended hospital birth as one of the two comparison groups.

The results of the evaluation found that FBC (n = 298) and Hospital (n = 2266) matched groups did not differ greatly in demographic, behavioral, and medical risk data. Differences that did emerge concerned relatively low percentages of women in either group. Comparison analysis across all three sites found no fetal deaths, very low proportions of complications such as Apgar scores less than seven in the first minute, low birth weight, and incidence of preterm birth (Table 2). The majority of outcomes were not statistically significant between groups. However, there were significantly fewer incidences of maternal complications found in two of the FBCs (0.23 and 0.03) than with their two hospital comparison groups (0.383 and 0.215). There were also significantly more infants transferred after birth at all three FBC sites as compared to their hospital comparison group (FBC: 0.029 v. Hospital: 0.0064; FBC: 0.044 v. Hospital: 0; FBC: 0.091 v. Hospital: 0.012). However, the comparison of rates of infant transfers is not an accurate interpretation of FBC relative safety because many hospitals have the necessary experts and equipment in-house for emergencies.

Table 2: Infant and Maternal Outcomes for the Kellogg Birth Center Evaluation (Cordray and Davidson, 2004)

	Urban birth center, Knoxville			**Urban birth center, Nashville**			**Rural birth center**		
Poor Childbirth Outcomes	FBC n=143	Hospital n=940	Odds Ratio	FBC n=90	Hospital n=1154	Odds Ratio	FBC n=33	Hospital n=172	Odds Ratio
Fetal death	0	0	--	0	0	--	0	0	--
Death < 29 days	0	0	--	0.011	0.001	0.08	0.030	0.006	0.2
Death > 29days < 1 year	0	0.001	--	0.011	0.001	0.08	0	0.006	--
APGAR <7 at 1 minute	0.028	0.069	2.6	0.011	0.047	4.4	0	0.047	--
APGAR < 7 at 5 minutes	0.007	0.005	0.8	0	0.009	--	0	0.023	--
Preterm (< 37 weeks)	0.021	0.055	2.7	0	0.030	--	0.030	0.064	2.2
Very Preterm (< 34 weeks)	0	0	--	0	0.003	--	0	0.006	--
LBW (< 2500 g)	0.014	0.019	1.38	0.033	0.036	1.1	0	0.029	--
VLBW (< 1500 g)	0	0	--	0	0	--	0	0.006	--
Infant transferred after birth	0.029	0.006	0.3	0.044	0	--	0.091	0.012	0.12*
Labor and delivery medical risks and complications	0.235	0.383	2.0**	0.157	0.18	1.14	0.030	0.215	8.8*

-- Not calculable; *p<0.05; **p<0.01

A total of three infants died after leaving the FBCs, and five infants died after leaving the comparison hospital groups. However, examination of information from death certificates did not indicate any direct linkage to actions taken or not taken by either FBC or hospital services. The evaluation concluded that the FBCs did not present additional problems for women and infants as compared to matched hospital groups.

Therefore, the FBC services were meeting their goal of providing safe prenatal and childbirth care.

In general, a review of the literature found that the majority of studies reporting maternal and infant outcomes at FBCs as compared to hospital groups report low rates of poor pregnancy outcomes. They also provide no indication that women are at greater risk for poor outcomes at FBCs than at hospitals (Cordray and Davidson, 2004; Jackson et al., 2003; Baruffi et al., 1984; Scupholme et al., 1986; Feldman & Hurst, 1987; Fullerton & Severino, 1992**;** Waldenstrom et al., 1997). However, very few studies have adequately addressed the known methodological concerns, such as designing adequate comparison groups and statistically controlling for biopsychosocial risk differences, which limit the use of current evidence in maternal health care systems.

To address these methodological concerns, the current analysis built on several methodological decisions made for the Kellogg Birth Center Evaluation and other previous work. As in the Jackson et al. (2003) study design, the current study also used low risk criteria to define the hospital comparison group and statistically adjusts for differences in risks between the two groups. The current study primarily is an extension of the Kellogg Birth Center Evaluation (Cordray and Davidson, 2004), which used the same sampling frame for constructing a second comparison group of midwife-attended hospital births. Furthermore, the current analysis expanded the evaluation by applying a propensity score adjustment to the comparison of FBC births and midwife-attended Hospital births. These methodological decisions were driven by current research on the potential sources of bias that exist in comparing childbirth services. A discussion of the

three primary sources of bias is needed to fully understand why it is necessary to control for these factors in observational studies.

Potential Sources of Bias in Research on Safety of Childbirth Services

Potential sources of bias in observational studies are created when treatment and control groups differ on important factors prior to treatment. Rosenbaum (2002) discusses the importance of considering both overt and hidden biases when planning an observational study. One critique in the FBC research literature is that women choosing FBCs are overall healthier than other women (Lieberman & Ryan, 1989; Albers & Katz, 1991), thus driving a selection bias that may distort any actual program differences. Three other important sources of bias in FBC research include: (1) differences in beliefs and attitudes driving women's choice of midwifery care over physician care; (2) differential use of technological interventions between midwives and physicians; and (3) differing levels of biopsychosocial risks known to affect pregnancy outcomes.

Potential Biases Related to the Choice of FBC and Midwifery Care.

Minimizing bias when investigating the safety of FBC services requires understanding factors that may influence a woman's choice for out-of-hospital care. There are a number of researchers that have found that beliefs about risk, use of technology, expectations of control, and information seeking behavior influence women's choice of provider and location for childbirth (Cohen, 1982; Fullerton, 1982; Galotti, Pierce, Reimer, & Luckner, 2000; Howell-White, 1999; Hundley, Milne, Glazener, & Mollison, 1997; van Der Hulst, van Teijlingen, Bonsel, Eskes, & Bleker, 2004). The

possibility that women who choose FBC care may differ on underlying psychological factors, as compared with women choosing hospital-based care, contributes to the source of self-selection bias.

Howell-White (1999) authored one of the few comparative studies assessing factors that influence both choice of provider and location, where all participants had knowledge of each option. Low medical risk women (N = 200) in this study had the option of choosing obstetrician care in a hospital (n = 102), nurse-midwife care in a hospital (n = 79), or nurse-midwife care in a FBC (n = 19). She found that factors influencing choice fell on a continuum. Factors that significantly predicted the probability of women's choice of obstetrician care over nurse-midwifery care included attitudes such as being less confidence in a nurse midwife (OR: 0 .347, $p<0.001$), expectations for less information about labor and delivery (OR: 0 .850, $p<0.01$), and higher perceptions of risk during childbirth (OR: 1.23, $p<0.001$). On the other hand, factors that significantly predicted the probability of women choosing nurse-midwife-FBC care over hospital care included higher levels of completed education (OR: 2.22, $p<0.05$), greater access to social support (OR: 1.13, $p<0.001$) and being less accepting of technological interventions (OR: 6.899, $p<0.001$). Profiles of women choosing nurse-midwife care in the hospital fell in the middle of the other two options. Significant predictors of choosing hospital care with a nurse-midwife over an obstetrician included having more confidence in a nurse midwife (OR: 2.66, $p<0.001$), having a greater expectation of information (OR: 1.18, $p<0.01$), and having less of a perception that childbirth is risky (OR: 0.835, $p<0.01$). In contrast, greater desire for the availability of technology should a complication occur (OR: 0.157, $p<0.01$) and greater access to social

support (OR: 0.921, $p < 0.05$) significantly predicted choosing a nurse-midwife in a hospital over a nurse-midwife at a FBC. The overall pattern of factors influencing choice provides evidence of how women who choose hospital versus FBC care, or women who choose physician versus nurse-midwife care may be different because of psychological factors.

The extent to which differential dispositional profiles related to choice influence pregnancy outcomes is an empirical question. Although there was no evidence found for this review that links belief about technology use, perceptions of risk, and confidence in provider as independently influencing outcomes, these factors could represent underlying coping styles. Women who have a positive coping style and see their pregnancy as controllable tend to choose better ways of coping with stress, and therefore, decrease its impact on their health during pregnancy (Dunkel-Schetter, Gurung, Lobel, & Wadhwa, 2001; Yali & Lobel, 1999). Social support is also believed to have both mediator and moderator effects on birth weight. Correlational studies consistently find main effects between prenatal social support and birth weight (Feldman, Dunkel-Schetter, Sandman, & Wadhwa, 2000; Norbeck & Tilden, 1983; Turner, Grindstaff, & Phillips, 1990). However, the main effect of social support and birth weight could also be evident because social support promotes positive health behaviors (Sheehan, 1998).

The possibility that psychological factors known to influence pregnancy outcomes may be connected to choice of service provides some evidence that these differences are important to consider when constructing comparable groups. Two studies, however, have found that women who actively search out midwifery care at a FBC do not differ on demographics, risk factors, or outcomes as compared to women who chose the FBC site

for other reasons, such as, location, finances, or overcrowding in hospitals (Jackson et al., 2003; Scupholme & Kamons, 1987). The findings from these two studies suggest that active choice may have no effect on outcomes. However, since ongoing research continues to suggest that psychological factors are important mechanisms influencing pregnancy outcomes, controlling for choice of service has the potential to minimize bias further when designing observational studies.

Potential Biases Related to Provider Practice

Differential use of technological interventions between obstetricians and nurse-midwives is well documented (Fullerton, Hollenbach, & Wingard, 1996; Oakley et al., 1996; Oakley et al., 1995; Paine et al., 2000; Rosenblatt et al., 1997), and could also be a potential source of bias when comparing outcomes between groups. While nurse-midwives typically practice in FBCs, physicians commonly provide care for most low risk women using hospital services. Researchers have found that for low medical risk women, nurse-midwives use 12.2% fewer resources than physicians including less continuous fetal monitoring and anesthesia, with no difference in Apgar scores, birth weight, or live births (Rosenblatt, et al., 1997). A randomized control trial of low risk women using hospital services also found that nurse-midwives used 9.4% less inductions, and were significantly less likely to use continuous fetal heart rate monitoring (Turnbull et al., 1996).

The advancement of obstetrical intervention has obviously been life saving to many women. However, there is much debate over whether there exist high rates of unnecessary interventions, especially for low medical risk women (Feldman & Hurst,

1987; Rooks, 1999; Wagner, 2001). In fact, the added risk of medical intervention, called iatrogenic risks (Wiener, 1998), is often referred to as the obstetrical cascade. The "obstetrical cascade" refers to the situation when use of one intervention increases the risk of needing a further intervention (Rooks, 1999). In addition to issues of choice, difference in provider behavior between midwives and obstetricians may also be a source of bias on important outcomes used to evaluate safety.

Potential Biases Related to Underlying Biopsychosocial Risk Profiles

Another potential set of biases concerns the possibility of differential risk profiles between groups that confound the treatment effect. The empirical literature widely agrees that there are multiple risks and multiple paths that determine poor pregnancy outcomes. However, many poor childbirth outcomes, such as low birth weight and preterm birth, are caused by a system of biopsychosocial risks with causal mechanisms that are not clearly understood (Berkowitz, 1981; Chomitz, Cheung, & Lieberman, 1995; Lobel, Dunkel-Schetter, & Scrimshaw, 1992; Paneth, 1995). Taking into account a system of biopsychosocial risks is necessary when constructing comparison groups for equating low risk status. This is evident in results from the Preterm Prediction Study (Iams et al., 2001) where, of the 2929 low risk women enrolled in the study, 127 (4.3%) had spontaneous preterm birth before 35 weeks gestation. Of these preterm births, 50% (64 cases) came from women without the additional risk factors of a previous preterm birth or spontaneous loss at less than 20 weeks gestation (Iams et al., 2001). It is obvious that unknown risks are part of the causal mechanisms of preterm birth.

In order to account for important biopsychosocial risks when assessing treatment effects, the current literature on risks of low birth weight and preterm birth was reviewed. Focusing on determinants of low birth weight and preterm birth help reveal the complexity of the biopsychosocial risk framework in pregnancy. Two major reviews identified 66 distinct risk factors that are associated with low birth weight and preterm birth (Institute of Medicine, 1985; Kramer, 1987). For the purpose of the current study, a summary of only the strongest risk factors is discussed. The biopsychosocial risk factors examined for the proposed study are organized in demographic risk markers, behavioral risks, biological risks predating pregnancy, and biological risks developed during pregnancy. The empirical information is taken from a small sample of studies in addition to the two larger reviews (Ickovics et al., 2000; Lieberman, Ryan, Monson, & Schoenbaum, 1987; Michielutte et al., 1992; Rauh, Andrews, & Garfinkel, 2001; Shiono, Rauh, Park, Lederman, & Zuskar, 1997; Zimmer-Gembeck & Helfand, 1996).

Demographic Risk Markers. Both low socioeconomic status (SES) and being an African-American woman are risks markers for low birth weight (LBW) and preterm delivery. Previous explanations have suggested that the lower SES groups experienced a high number of life stressors, high levels of psychological distress, and were more at risk for poor health behaviors, lived and worked in risky environments (Taylor, Repetti, & Seeman, 1997), and had less social support and other resources to protect against stress (Kramer, Seguin, Lydon, & Goulet, 2000; Rini, Dunkel-Schetter, Wadhwa, & Sandman, 1999). The theories trying to address racial disparities considered risks, such as lack of appropriate cultural support (James, 1993), discriminatory contact leading to chronic stress (Wise, 2003), and a "weathering" hypothesis that takes into account the life-time

effects of living in poverty (Geronimus, 1996). For example, Rauh et al. (2001) suggests that the bulk of the racial differences between the risk of older maternal age affecting LBW is driven by individual poverty. Rauh et al. (2001) found that older, African-American women receiving Medicaid had significantly greater odds of predicting LBW than those women not receiving Medicaid.

Consistent with studies assessing the effect of race, Lieberman et al. (1987) found that African American women are approximately two times as likely to deliver a premature (< 37 week) baby than White women (OR: 1.94, $p < .05$) (Lieberman, et al., 1987). Lieberman et al. accounted for all of the race effect by controlling for a 1-point hematocrit change (a physiological marker for anemia) and an aggregated education, demographic, and behavioral variable (EDB). The EDB variable assumes that age < 20; single marital status; no high school education; and receiving welfare support are interchangeable. Combination effects were scored as having 0, 1, or 2+ EDB risk factors during pregnancy. Nearly all of the increased risk of premature delivery for African American women was accounted for when EDB and hematocrit levels adjusted for African American race (OR= 1.03, NS).

Although their primary focus was on accounting for the effect of race, Lieberman et al's (1987) study reported information useful to evaluating the interrelationships between other risks. Several medical and behavioral risk factors were also significantly related to the number of EDB risk factors. In the full sample, rates of cigarette smoking, urinary track infection, malformed infants, ponderal index, and low hematocrit levels all significantly increased ($p < .05$) with additional EDB risk factors. The odds for premature

delivery also increased as women were exposed to more EDB risks (Lieberman et al., 1987).

Behavioral Risk Factors. The most consistent and direct predictive risk for low birth weight is cigarette smoking (OR range from 1.68 to 3.17) (Kramer, 1987; IOM, 1985; Zimmer-Gembeck & Helfand, 1996; Michielutte et al., 1992; Rauh et al., 2001). Kramer (1987) also listed consumption of more than two drinks of alcohol per day as increasing the likelihood of low birth weight (RR: 1.78). Odds of low birth weight are also at least two times greater for substance users (Rauh et al., 2001). Nutritional risks such as low pre-pregnancy weight (RR: 1.84), and poor gestational weight gain (RR: 1.98) are also important behavioral risks that are known to increase the risks of low birth weight (Kramer, 1987).

Biological Risks Predating and Developing During Pregnancy. Obstetric history, including parity (a first birth), previous preterm or low birth weight infant, and previous terminations, significantly impacts the likelihood of problems in the current pregnancy (Zimmer-Gembeck & Helfand, 1996; Michielutte et al, 1992). The Institute of Medicine (1985) reviewed a large number of risks that were thought to predict low birth weight and preterm birth. The IOM committee reported the average relative risk ratios (i.e. the proportion of risk in one group divided by the proportion of risk in a second group) found in the studies they reviewed. Women with diabetes (RR: 5.5) and hypertension (RR: 6.2-40.4) are considered at most risk for poor pregnancy outcomes (IOM, 1985; Kramer, 1987). Additional medical risks, such as having twins (RR: 2.0-5.5), hypertension with pre-eclampsia (RR: 2.4-5.8), 1st or 2nd trimester bleeding (RR: 2.1), placental problems, such as abrupto placenta or placenta previa (RR: 8.0), fetal anomalies (RR: 2.4),

spontaneous premature rupture (RR: 2.0), oligohydramnios (RR: 2.6), hypotension (RR: 1.9-4.2), anemia (RR: 2.1-4.2), and isoimmunization (RR: 4.3) all have greater than a 2.0 relative risk for incidence of low birth weight and premature birth (IOM, 1985). As reported by other studies, these risks are likely to have the largest magnitudes of effect on LBW and preterm birth over any other risk factor (IOM, 1985; Zimmer-Gembeck & Helfand, 1996). However, they are also likely to have very low prevalence in the population. Zimmer-Gembeck and Helfand (1996) estimate that less than 2% of women experience bleeding and other medical risks from their aggregated index. Even with their low prevalence, these risks should be included as part of the causal system of risks leading to LBW. Inclusion of these risks will likely account for some variance in birth weight.

Exposure to more than one of these risks also increases the likelihood for poor outcomes. Zimmer-Gembeck and Helfand (1996) used a medical composite score that included hypertension, prior renal disease, diabetes, oligohydramnios, urine protein, or structural abnormality of the cervix. They found that exposure to one of these risks significantly increased the odds of low birth weight (OR: 2.40, 95%CI: 1.56-3.71); however, having more than one of these risks was associated with an even larger increase in risk (OR: 7.03, 95%CI: 2.18-22.72).

Psychological Risk Factors. Research into the specific roles of stress during pregnancy has increased due to two principal factors: First, research suggests that only a half to three-fourths of the cases of adverse birth outcomes can be accounted for by biological risk factors (Shiono & Behrman, 1995). Second, research also suggests that studying mechanisms of stress is one way to understand the linkages between SES,

ethnicity, and health (Stanton, Lobel, Sears, & DeLuca, 2002). Consistently, reviews find that psychological factors act as mediators and moderators that either protect from or promote a cascade of events related to poor reproductive health outcomes. There is empirical agreement suggesting that stress can impact pregnancy outcomes with direct paths through physiology, and indirect paths through lifestyle and behavior (Chomitz, Cheung, & Lieberman, 1995; Dunkel-Schetter et al., 2001; McAnarney & Stevens-Simon, 1990; Paarlberg, Vingerhoets, Passchier, Dekker, & Van Geijn, 1995; Stanton et al., 2002). Although the current study does not measure stress directly, it is important to point out that the unmeasured psychological factors may influence outcomes between FBC and hospital services.

Any differences between comparison groups in the set of biopsychosocial risk factors that are known to influence poor pregnancy outcomes can become a source of bias limiting the interpretation of treatment effects. Since a common critique of previous studies has been that women self-selecting FBCs are overall healthier than women in the hospital (Albers & Katz, 1991; IOM, 1982; Lieberman & Ryan, 1989), it is important to be clear about the differences between complete risk profiles before testing for differences between FBC and hospital services. The current study combines the knowledge of important biopsychosocial risk factors with what is known about factors that influence women's choice of service, and employs better strategies based on empirically driven criteria to further minimize bias in observational studies and strengthen the quality of evidence evaluating the safety of FBC services.

CHAPTER III

METHODS

The current study evaluated the safety of FBCs by employing methodological strategies that reduce sensitivity to hidden biases common in observational studies with nonrandomized comparison groups (Little & Rubin, 2000; Rosenbaum, 2002). Comparing maternal and infant outcomes between FBCs and hospital-based services adds to the current evidence evaluating whether FBCs can effectively intervene on risks in pregnancy and childbirth as compared with hospital services. As presented in the literature review, two primary limitations challenge the validity of current evidence: (1) the lack of adequate comparison groups to minimize the effect of self-selection bias, and (2) the lack of adequate sample sizes necessary for detecting effects of treatment when outcomes are rare (Albers & Katz, 1991; Fullerton & Wingard, 1990; Jackson et al., 1994; Lieberman & Ryan, 1989). To address these critiques, methodological strategies for this study included constructing comparison groups that closely match characteristics linked with FBC care, applying statistical adjustments to minimize the effects of hidden biases, and using a large sample of birth certificate records from the State of Tennessee.

As suggested by previous work in the design of observational studies (Rosenbaum, 2002; Little & Rubin, 2000), developing appropriate comparison groups requires that the sampling frame include adequate matches on personal characteristics and, in this case, risk profiles. Without comparable risks, results are vulnerable to selection bias. The methodological concern in previous FBC research was that women

receiving care at FBCs are overall healthier than the women in hospital comparison groups (Jackson et al., 2003; Lieberman & Ryan, 1989). This concern implied that non-significant differences in poor outcomes at FBCs and hospitals are a reflection of the case-mix, not the treatment effect. In order to study the treatment effect, comparison groups must have comparable levels of risk.

Only the low medical risk population of childbearing women is eligible for FBC care (NACC, 1999). Challenges to creating comparable groups in FBC research include: (1) different beliefs and attitudes driving women's choice of midwifery care at FBCs from women choosing physician care in hospitals (Howell-White, 1999; Oakley et al., 1996; Oakley et al., 1995; Rosenblatt et al., 1997); (2) differences in provider practice between obstetricians and nurse-midwives (Fullerton et al., 1996; Oakley et al., 1996); and (3) variable childbirth outcomes, such as low birth weight and preterm birth, which may be caused by a system of biopsychosocial risks with mechanisms that are not clearly understood (Paneth, 1995; Stanton et al., 2002). By combining what is empirically known about differences between FBC and hospital services with what is known about risk factors that impact poor pregnancy outcomes, appropriate comparison groups and statistical adjustments were employed to help minimize the effect of hidden biases on measuring treatment effects.

The current study is an extension of the Kellogg Birth Center Evaluation Project (Cordray & Davidson, 2004) that found no problematic maternal and infant childbirth outcomes more likely to occur in any of the three FBCs as compared with hospital groups in Tennessee from 1999-2002. This study used the same comparison group that the Kellogg evaluation used, which is a group of midwife-attended hospital births to control

for similar characteristics of women choosing midwives as providers in FBC care. However, this study expanded the number of years examined to twelve, from 1990-2002, to provide a sample size large enough to detect small effects of treatment. The current study also used logistic regression to model the initial group differences, created a propensity score for each case and then used sub-classification analysis on the propensity score to determine if group differences were properly balanced by the propensity score analysis. This approach has not been used previously in studies investigating the effectiveness of FBC services.

The analytical plan consisted of two separate analyses using different comparison groups. In the first analysis, outcomes of FBC births were compared to low medical risk hospital births in East Tennessee. Results from this comparison reflected previously designed studies, and therefore, were used as a benchmark to compare the outcomes from Tennessee birth centers with results from previous research (Jackson et al., 2003). Following Corday and Davidson (2004), in the second analysis, outcomes for FBC births were compared with births attended by midwives in hospital settings. In addition to controlling for geographic area, this analytical strategy controlled for differences in philosophy of care between physicians and nurse-midwives. Both sets of analyses controlled for the differential risk profiles between FBC and the appropriate hospital group (Low Risk or Midwife Attended). In addition to equating groups on service choice and procedures, the current analyses went beyond those analyses performed by Cordray and Davidson (2004) by explicitly modeling, through the use of propensity analysis, any remaining risk factors associated with the choice of FBC or Midwife-based services.

Hypotheses

Studies reviewed for this study consistently reported that the rates of morbidity and mortality outcomes were statistically non-significant between FBC and hospital services for low risk pregnant women and for women choosing a midwife to attend their birth within a hospital setting. Therefore, hypotheses were as follows for each analysis:

Hypothesis 1: Women experience equivalent rates of complications during childbirth at FBCs and hospital services.

Hypothesis 2: Infant morbidity rates are equivalent at FBCs and hospital services.

Because of the closer matching of service-type, it was expected that infant and maternal outcomes would be more "equivalent" for women in the FBC versus midwife-attended Hospital groups than women in the FBC versus low risk Hospital groups.

Secondary Data Source

Data were obtained from the Tennessee Department of Health Statistics[1]. The majority of risk and outcome data were collected from the complete set of birth certificate records maintained by the State of Tennessee from 1990 through 2002. Death certificates (indicating a child died within one year of birth) and fetal death certificates (indicating a stillbirth after 28 weeks gestation) were linked with the birth certificates to obtain measures of mortality as suggested by previous research on childbirth outcomes (Buehler, Prager, & Hogue, 2000; Gould, 1999).

[1] As this study is an extension of a longitudinal evaluation on freestanding birth centers in TN, the Kellogg Birth Center Evaluation Project, permission to use the secondary data on birth certificates to evaluate safety of FBC services was previously approved by both Vanderbilt University (#000563) and State of Tennessee Institutional Review Board requirements for studying human subjects.

Sampling Frame

The records obtained for this study represent a sample of women receiving childbirth services from freestanding birth centers and two different samples of women receiving hospital services in Tennessee.

Freestanding Birth Center (FBC) Group

The *FBC* group (n = 2,463) consisted of all births occurring at two eastern Tennessee freestanding birth centers from 1990 through 2002. One center is located in an urban area (n = 1104) and the other center is located in a rural area (n = 1359) of this region. Although there are six freestanding birth centers in Tennessee, this research was restricted to two mature FBCs that have at least a 10-year operational history, therefore providing a sizable sample and suggesting organizational stability. Organizational stability helps rule out any influence of administrative start up factors on health outcomes.

The data recorded on birth certificates indicated the location of birth, and if applicable, the facility a woman was transferred from if more extensive intervention was needed during childbirth. As prescribed by an intent-to-treat model, the FBC group included both women who delivered at the two FBC sites (n = 2240) and those who were transferred from a FBC during childbirth (n = 223)[2].

[2] The sample of transfer cases was an underestimation of the total number of women who transferred from FBC care for medical or personal reasons during prenatal care or labor. The sample also did not represent women who started prenatal care at a FBC and whose pregnancy survived for less than 28 weeks gestation. Although these are important aspects of evaluating the relative safety of this service, the access to only birth certificate data did not allow for a more precise sample. The limitation section of the Discussion, Chapter V, discusses this issue in more detail.

Hospital-based Comparison Groups

Both comparisons maximized geographical similarities with the FBC group, thereby controlling for unobserved community and cultural factors that may impact outcomes. Geographical differences, such as access to care, community beliefs, and environmental exposures are known to influence health behaviors and pregnancy outcomes (Buka, Brennan, Rich-Edwards, Raudenbush, & Earls, 2003; English et al., 2003; Taylor et al., 1997). Therefore, the sampling frame was limited to five eastern Tennessee counties and years that match where the birth certificate indicated the use of FBCs.

To create comparison groups based on similar geographical residence, the pool of midwife-attended hospital births was first limited by only those births that occurred in the same counties as the two FBCs used in this study. This limit produced an insufficient number of records; therefore, the inclusion criteria were expanded to include midwife-attended hospital births in surrounding counties where FBC clients reside. In addition to geography, comparison groups were also matched on the comparable years for each FBC site. For the urban FBC comparison group, records were limited to midwife-attended hospital births in the counties of Knox and Hamilton from 1991 through 2002; and for the rural FBC comparison group, records were limited to the counties of Monroe, McMinn, and Bradley from 1990 through 2002.

FBC versus Low Risk Hospital Births (FBC v. Low). The first analysis compared FBC births with a sample of low-risk hospital births in corresponding geographical areas (FBC-Low). The comparison group consisted of a random sample of low medical risk women using hospital services in the five selected counties in eastern Tennessee (n =

2,500)[3]. This sample represents twelve hospital facilities. The structure of the birth certificate data did not allow use of a comprehensive low risk index for implementing exclusion criteria because of its lack of specificity to certain variables. Therefore, criteria for defining low-risk women were taken from three studies assessing the effects of electronic fetal monitoring (Amato, 1977; Eden, Seifert, Winegar, & Spellacy, 1987; Leveno, Cunningham, & Nelson, 1986). Exclusion criteria for classifying a woman as a "low risk" pregnancy were women younger than 16 and older than 39, occurrence of twins or multiples, any congenital anomalies, postdate > 42 weeks gestation, diabetes, hypertension and preeclampsia, vaginal bleeding, cardiac or lung disease, oligohydramnios, anemia, renal disease, incompetent cervix, and Rh sensitization[4].

The National Birth Center Study (Rooks et al., 1992a, 1992b, 1992c; Rooks et al., 1987) used results from these three studies: Amato, 1977; Eden et al., 1987; Leveno et al., 1986 to compare their FBC outcomes of mortality, Apgar scores, and cesarean section rates. Although this method of comparison was criticized because of the use of different data sources, this study used the same low risk criteria. The rationale for using a low risk control group was to provide a base line for interpreting outcomes based on criteria used in previous studies of outcomes in low risk pregnancies. Furthermore, this comparison provided a baseline for comparing results in the second comparison group, midwife-attended Hospital births. Overall, the design of this analysis used similar logic as

[3] There were 150,506 low risk hospital births in the sampling region of five eastern Tennessee counties. A random sample of 2,500 cases was chosen from this group using the SAS system PROC SURVEYSELECT. All proportions of risks, and later outcomes, were checked against the full set of 150,506 cases to assure that the random sample was representative of the population of low risk in these counties. In the random sample, there was no more than a 0.05 difference in proportions of risks and no more than a 0.01 difference in outcomes than seen in the full population.

[4] These criteria are consistent with definitions of low risk pregnancy found in two relatively recent studies using low-risk populations to assess cesarean delivery rates (Gould et al., 2004; Seshadri & Mukherjee, 2005).

previous studies for evaluating the safety of FBCs (Fullerton, et al., 1993; Rooks et al., 1989).

FBC versus Midwife-Attended Hospital Births (FBC v. Midwife). The comparison group for the second analysis (FBC-Midwife) included a sample of records from women with a midwife-attended hospital birth in similar geographic areas as women included in the FBC group (n = 11,635). This sample represented midwife-attended hospital births from seven different hospital facilities. There were no additional exclusion criteria based on medical risks, because it was assumed that midwives in hospitals attend primarily low-risk births. The goal of using criteria that are more restrictive to construct the comparison group for the FBC-Midwife analysis was to minimize bias using empirically driven criteria. Constructing the comparison group by using empirically driven strategies for controlling effects of provider and location helped control for any unobserved factors influencing group assignment and childbirth outcomes. Limiting providers in a comparison group to nurse-midwives controlled for known, but unobserved covariates that could contribute to an effect of selection bias. These unobserved covariates included factors connected with practice differences between obstetricians, family practitioners, and nurse-midwives in the use of obstetrical intervention (Hundley et al., 1994; Oakley et al., 1996; Rosenblatt et al., 1997; Turnbull et al., 1996) and represented the variability in beliefs and attitudes of women who choose between physicians and midwives (Howell-White, 1999; Oakley et al., 1996).

Power Analysis[5]

The sample size in the FBC-Low analysis provided statistical power of at least 80% ($\alpha = 0.05$) to detect a significant odds ratio as small as 0.8 or 1.2 if the proportion of poor outcomes is as low as 0.20 in the Hospital control group. The sample size in the FBC-Midwife analysis provided statistical power of at least 80% ($\alpha = 0.05$) to detect a significant odds ratio as small as 0.8 or 1.2 if the proportion of poor outcomes is as low as 0.10 in the Hospital control group.

Measurement

The variables used in this study represent a subset of risks and outcomes that have been previously identified as important indicators of poor pregnancy outcomes (Kramer, 1987; Paneth, 1995), and are available on the official Tennessee birth certificate.

Group Membership

Group membership was dichotomously coded as FBC group =1 and Hospital group = 0, using the sampling frames explained previously.

Biopsychosocial Risks

Risk variables were clustered together based on the biopsychosocial framework of risk (Table 3)[6]. Biopsychosocial risks were dummy coded into dichotomous variables (1

[5] Calculating power, when using an odds ratios statistic, is dependent on the proportion size of the outcome in the control group. For this study, achieving adequate power with large sample sizes is deceptive because the proportions of poor outcomes in low risk pregnancies are very small. To find statistical differences that are very small requires very large sample sizes.

[6] Detailed descriptions of risk variables are summarized in Appendix A, Table A-1.

= risk is present, 0 = risk is not present). Incidence of medical risk is an aggregated factor consisting of 15 individual medical risks. The rationale for creating an aggregated variable was that eligibility criteria for designing the low risk Hospital group eliminates 10 of the 15 medical risks, and for all risks the proportions in the FBC group were small (Appendix Table A-2). If a woman had an incidence of any one of the 15 medical risks, that case was coded as 1. If there were no incidences of any medical risks, that case was coded as 0[7].

Table 3: Biopsychosocial Risks Included in the Risk Profile

Demographic risk markers	**Behavioral risk factors**	**Biological risk factors**	
Age Risk <19 years old; Age Risk >35 years old; Education < 12 grade; Maternal race (African American); Unmarried status	Entering prenatal care >20 weeks gestation; Cigarette smoking; Alcohol use; Drug use	Nulliparous; Previous infant 4000+ grams; Prior history of small-for-gestational age; Prior history of preterm birth; Greater than 2 previous pregnancy terminations; Other risks (general risk category)*	Incidence of medical risk^

* Other medical risk factors experienced by the mother that may cause or contribute to complications of this pregnancy. Examples are AIDS, preeclampsia, rubella, syphilis, gonorrhea, early onset of delivery and mental disorder.

^ The aggregated variable for medical risks includes incidence of anemia, cardiac disease, acute or chronic lung disease, diabetes, active genital herpes, hemoglobinopathy, chronic hypertension, renal disease, hydramnios or polyhydramnios, oligohydramnios, hypertension, pregnancy-induced, eclampsia, incompetent cervix, Rh sensitization, and uterine bleeding.

Outcome Variables

Important maternal and infant morbidity outcomes cited in previous literature (Jackson, et al., 2003; Rooks, et al., 1989) were included as primary outcomes for

[7] A count of medical risks for each case revealed that only 15 cases in the FBC group had incidence of two risks. 91% of all cases had no medical risks and 8.4% of cases had one risk. In the low risk hospital comparison group only 11 cases had two risks, and in the midwife-attended hospital comparison group only 246 (2.1%) of cases had more than one risk.

estimating safety of FBC services in comparison to hospital services (Table 4). Each outcome was dummy coded as a dichotomous variable (1 = outcome is present, 0 = outcome is not present). Two outcomes, abnormal conditions for the infant and maternal complications in labor, were aggregated variables. It is common practice to aggregate sets of outcomes that occur very rarely in low-risk populations (Jackson et al., 2003; Villar et al., 2001)[8]. The aggregated risks of abnormal conditions for the infant and maternal complications were coded: 1, incidents of complications, and 0, no incidents of complications. Detailed descriptions of each outcome are listed in Appendix Table A-3.

Table 4: Maternal and Infant Outcomes

Infant Outcomes	**Maternal Complications**
Infant mortality, fetal and infant death	Incidence of maternal complications during labor ^
Low birth weight, < 2500 grams	
Small for gestation age	
Large for gestation age	
Preterm birth, <37 weeks gestation	
Apgar score, < 7 at 1 minute	
Apgar score, <7 at 5 minutes	
Fetal distress	
Incidence of abnormal conditions*	

* The aggregated variable for incidences of abnormal conditions for the infant include: anemia, birth injury, fetal alcohol syndrome, fetal drug syndrome, hyaline membrane disease/RDS, maconuim aspiration syndrome, assisted ventilation < 30 minutes, assisted ventilation > 30 minutes, seizures, and a general other category.

^ The aggregated variable for incidence of maternal complications during labor includes: febril, >100 degrees Fahrenheit, moderate/heavy meconium, premature rupture of membranes (>12 hours and < 24 hours), prolonged rupture of membranes (>24 hours), abruptio placenta, placenta previa, other excessive bleeding, seizures during labor, precipitous labor (< 3 hours), prolonged labor (> 20 hours), dysfunctional labor, breech, malpresentation, cephalepelvic disproportion, cord prolapse, anesthetic complications, and a general other category.

Although mortality rates are prevailing indicators of safety, infant and maternal mortality are very rare in studies of low risk pregnancies. Therefore, most studies cannot

[8] Nevertheless, the frequencies of cases having more than one of the complications that are part of the aggregated variable were checked to determine whether this practice was relevant to this data set. In the abnormal conditions of the infant variable, only 22 cases (0.009) in the FBC group had more than one incidence. And in the maternal complication variable, only 101 cases (0.041) had more than one incidence.

meaningfully analyze group differences. This study expected to find similar rates of mortality in the FBC group as found in the Kellogg evaluation results (Cordray & Davidson, 2004). Therefore, consistent with the decisions made in a Cochrane systematic review (Olsen & Jewell, 2003) and other studies of low risk pregnancies (Hundley, et al., 1994), only the proportion of infant mortality was compared between groups. An index of maternal mortality is not available on the birth certificate.

Reliability of Birth Certificate Data

Research has shown that the reliability of birth certificate data differs widely among variables causing the use of birth certificate records to be controversial in the study of obstetric and perinatal outcomes (Buescher, Smith, Holliday, & Levine, 1987; DiGiuseppe, Aron, Ranbom, Harper, & Rosenthal, 2002; Dobie et al., 1998; Gould, 1999; Piper et al., 1993)[9]. To obtain measures of reliability, birth certificate data have been compared with information abstracted directly from the medical chart, the industry "gold standard". Studies reviewed generally agree that demographic (i.e., zip codes, race, marital status, age, and education) and primary infant outcome (i.e. birthweight, gestational age, and apgar scores) variables have high rates of agreement between medical records and birth certificates. In contrast, risks and complications that rarely occur in the population have low to medium agreement and are shown to be an underestimation of the incident rate found in the population (Piper et al., 1993;

[9] Three studies were selected to provide a general estimate of birth certificate reliability based on their relevancy to the current study. Piper et al. (1993) examined certificates from Tennessee and calculated reliabilities for a very low birth weight (<1500 grams) group, as well as a randomly sampled group of normal weight infants. Dobie et al. (1998) measured reliabilities for low medical risk women in Oregon, and DiGiuseppe et al. (2002) compared reliability rates between teaching and non-teaching hospitals in Ohio. Buescher et al (1993) evaluated a general sample of 1989 North Carolina birth certificates as compared to hospital records.

DiGiuseppe et al., 2002; Dobie et al., 1998). Since the purpose of this study is to assess the relative differences between FBCs and hospital services, it is expected that any problems with reliability will be equally distributed across the groups.

Despite the inconsistencies, the use of birth certificates continues to be an important source of information for providing estimates of national trends and outcomes, and its use is encouraged when limitations are addressed (Gould, 1999; Ventura, Hamilton, Mathews, & Chandra, 2003). For example, a study by the Center for Disease Control found that smoking during pregnancy, which is thought to be frequently underreported, correctly reflects the trends and variations nation-wide (Ventura et al., 2003). In this study, several of the infant outcomes used as markers for FBC safety have relatively high reliability coefficients (i.e. birth weight, gestation age, and Apgar scores). The obvious benefit for using birth certificates as the data source in this study was the ability to obtain a large sample size over time.

Missing Data

Variables were surprisingly complete in the current dataset. The specific number of missing cases in each analysis is documented in the results section. Research in perinatal outcomes using vital records suggests that cases with missing data are not randomly distributed, but are concentrated in socially at-risk groups that have higher rates of infant mortality and morbidity (Gould, et al. 2004; Gould, 1999). For this reason, Gould (1999) encourages dropping cases with missing data from any risk-adjustment analyses. After examining all risks and outcomes in both FBC-Low and FBC-Midwife comparison, the amount of missing data was minimal. Specific proportions of missing

data are reported with the results. Because of the low rates of many conditions and outcomes, imputation was not attempted.

Analysis Plan

The analytical plan consisted of two primary steps. The first step was to evaluate risk profiles between FBC and Hospital groups to reveal pre-existing differences. The second step was to use statistical adjustment techniques to control for pre-existing differences. These two steps were approached in two different ways. For the logistic regression-based analyses, the univariate differences were examined in step 1. These risk factors were then included in the logistic regression as control variables (along with the group designation). In the propensity-based analysis, the first step entailed constructing a linear composite (the propensity score) to represent the differences between groups; in the second step, the propensity analysis was used as the control variable in a logistic regression that also includes the group designation. The propensity score analysis was expected minimize imbalance of covariates (across groups); in turn, it was expected to minimize bias in the treatment effect.

In order to interpret the safety of FBC services as compared to hospital services on all outcomes, beta coefficients calculated using logistical regression techniques[10] were converted to odds ratios and 95% confidence intervals so that the direction and magnitude of effects were more interpretable. Odds ratios were interpreted in this study as the likelihood of one group to have an incidence of an outcome as compared to the other group. Odds ratios are considered the most meaningful statistic for assessing the

[10] A logit probability model is used for estimating all dichotomous outcomes. A check for multicollinearity was conducted and any effect of high correlations was ruled out (Tabachnick & Fidell, 2001).

relative risk in complications that are fairly rare (Motulsky, 1995). To calculate the odds ratio, the number of incidences of an outcome is divided by the number of non-incidences of an outcome within groups. This produces the odds for each group. Then the odds of occurrence in the FBC are divided by the odds of occurrence in the Hospital to produce the ratio.

The FBC v. Low Risk Pregnancy Comparison

The first comparison between FBC and low risk Hospital groups (FBC-Low) and the second comparison between FBC and midwife-attended Hospital groups (FBC-Midwife) used parallel steps in their analyses. First, frequencies, proportions, and odds ratios were calculated for each risk between FBC and Hospital groups. This information helped evaluate the extent to which risk profiles are different between groups. Next, the frequencies and proportions of infant and maternal outcomes were examined individually and an unadjusted odds ratio and 95% confidence interval were calculated to characterize the extent of the differences between groups. The third step used a multivariate logistic regression model to calculate an adjusted odds ratio and 95% confidence interval for each outcome using all demographic, behavioral, and medical factors in the risk profile as control variables. Theoretically, by including all risks in the model, the outcomes were controlled for differences on the risk profile. The initial model equation for using a multiple logistic regression model to calculate odds ratio was as follows: $\Pr(Y_i=1) = 1/(1+\exp^{-Z})$; Where $Z = b_0 + b_1X_1 + b_2X_2 + \ldots + b_{16}X_{16}$; Y = Outcome; X_1 = group; X_2 = young age; X_3 = old age; X_4 = race; X_5 = education risk;X_6 = unmarried; X_7 = nulliparous; X_8 = cigarette smoking; X_9 = alcohol use; X_{10} = drug use; X_{11} = previous

small for gestational age infant; X_{12} = previous infant greater than 4000 grams; X_{13} = more than two previous pregnancy terminations; X_{14} = previous preterm infant; X_{15} = incidence of medical risks; X_{16} = incidence of generalized other medical risk.

Beyond Observed Variables: The FBC v. Midwife Comparison

The FBC-Midwife comparison was further examined using an additional statistical adjustment and sub-classification on the propensity score (Rubin, 1997; Rosenbaum & Rubin, 1984). Although controlling for risks by using multivariate logistic regression models is a conventional approach, statistical adjustments using propensity score technology have been shown to be a good option for balancing risk distributions between nonrandomized groups with large number of covariates (Dehejia & Wahba, 2002; Rosenbaum & Rubin, 1984). Before continuing, a brief explanation of theory and procedures is necessary for calculating the propensity score.

Use of Propensity Scores

The economic and evaluation literatures have discussed propensity score technology as a promising procedure for controlling confounding of causal effects in observational research designs (Dehejia & Wahba, 2002; Rosenbaum & Rubin, 1984; Rubin & Thomas, 1996; Winship & Morgan, 1999). A propensity score, $\lambda(x)$, is the conditional probability that an observation receives treatment versus a control given a group of pretreatment covariates. Employing the procedure entails a two-step process. First, the propensity score is estimated using a standard probability model. Second, the propensity score, itself, is used to control for confounding or as an index for matching

cases from a control group. Sub-classification on the propensity score is a procedure for stratifying cases based on their probability of receiving treatment (Rubin, 1997; Rosenbaum & Rubin, 1984). Successful adjustment on the propensity score theoretically leads to a balanced design where group assignment into a treatment and control condition does not differ with respect to any of the observed covariates used in calculating the propensity score (Dehejia & Wahba, 1999; Rosenbaum & Rubin, 1984; Winship & Morgan, 1999).

There are several theoretical benefits to using a propensity analysis over standard regression models. First, the groups are balanced with a single dimension, Z_i, rather than multiple dimensions representing each variable (Rosenbaum & Rubin, 1984; Rubin, 1997; Winship & Morgan, 1999). Therefore, evaluating the group distributions of propensity scores allows for a more straightforward assessment of the amount of overlap between groups on the joint distribution of the covariates. This is especially useful when there are many covariates, such as in the current study (Rubin, 1997). Second, if propensity scores are split into five equal subclasses and successfully balanced within each sub-classes, the theorem of sub-classification states that stratification in to quintiles removes approximately 90% of the bias based on the observed covariates (Cochran, 1968). Third, this process also does not make a distinction between variables that are highly or weakly predictive of outcomes (Rubin & Thomas, 1996). Therefore, propensity score technology is thought to provide a closer estimation of a treatment effect obtained from a randomized experiment than other methods of standard non-experimental estimators (Rubin, 1997; Rosenbaum & Rubin, 1984). Fourth, since the probability model used to estimate the propensity score does not take into account the outcome variables in any form, no

behavioral assumptions are attached. Therefore, this procedure is not considered data mining (Dehejia, 2005). The process serves only to reduce the dimensions of the multiple covariate risk factors into one dimension, thereby making treatment assignment ignorable.

There are also limitations to this process (Rubin, 1997; Rosenbaum & Rubin, 1984), which have been shown empirically to restrict the number of contexts that are appropriate for its use (Austin, Mamdani, Stukel, Anderson, & Tu, 2005; Cepeda, Boston, Farrar, & Strom, 2003; Michalopoulos, Bloom, & Hill, 2004; Winkelmayer & Kurth, 2004). First, unlike randomization, the procedure cannot balance the distributions of unobserved covariates. Any unobserved, yet influential variables on group assignment are not included, and therefore they remain a source of bias. Second, covariates that are affected by the exposure of interest cannot be used in the propensity model. And third, studies have found that the theoretical advantage of using propensity scores to adjust for confounding does not have any great advantage over traditional regression models unless the number of outcomes observed is a ratio of seven or less, per covariate (Cepeda, et al., 2003)[11]. In studies with rare outcomes relative to the number of covariates and large sample sizes, however, use of propensity scores plays a legitimate role to help control for confounding (Winkelmayer & Kurth, 2004; Cepeda, et al., 2003).

Sub-classification on the Propensity Scores. Considering these limitations in the context of the current study, sub-classification on the propensity score has the potential to play a legitimate role in minimizing additional bias in comparing treatment effect between FBCs and Hospital groups. By restricting one of the comparison groups to only

[11] This limitation is consistent with other work using Monte Carlo studies to investigate bias in logistic regression models. When studies have less than 10 outcomes per covariate, there is an increase in the effect of bias (Peduzzi, Concato, Kemper, Holford, & Feinstein, 1996).

midwife-attended births, there is potential that some of the unobserved biases based on provider practice and patient preferences are controlled. Also, there are no indications that prenatal care affects the presence of any of the covariate risks used in this model. Even risks that present themselves during pregnancy most likely have some underlying mechanism that predisposes women to manifesting risk. Propensity scores have not been used to statistically balance the multiple risk factors known to influence pregnancy outcomes. Therefore, utilizing propensity score technology for estimating safety of FBCs, as compared to hospital childbirth services, has the potential to provide better estimates of treatment effect than those obtained in previous research by reducing any effect of self selection bias.

Procedures for Estimating the Propensity Score. Calculating the propensity score requires an iterative procedure that continually checks for an appropriate fit with the entire set of observed covariates (Rosenbaum & Rubin, 1984). Achieving balance requires estimating propensity scores, examining distributions of propensity scores, stratifying the scores into subclasses, and testing for statistically significant differences in the distribution of covariates between FBCs and Hospital groups within each subclass.

The following analytic procedures were adopted from suggestions made by both pioneers of the technique and more recent adaptors (Dehejia & Wahba, 2002; Rubin, 2001; Rosenbaum & Rubin, 1984; Rubin, 1997; Winship & Morgan, 1999).

(1) Frequencies and odds ratios for all risk variables were calculated to provide base-line differences between groups[12].

[12] Refer to Table 5.

(2) The propensity score for each case was estimated using a logistic regression to find the logit probability model[13]: Pr(*Ti*=1 given Xi) = $\exp^{\lambda h(Xi)}/(1+\exp^{\lambda h(Xi)})$, where Ti = treatment status and h(Xi) is the list of all biopsychosocial risks from Table 3.

(3) Criteria were checked for evaluating the extent to which the two groups have propensity score distributions that overlap. Given the large number of covariates included in the risk profile, Rubin (2001) sets criteria that help distinguish whether the set of covariates in the two observational groups overlap enough for a regression adjustment to be "trustworthy". The criteria included: (a) that the difference in means of the propensity scores between groups are small (i.e. less than half a standard deviation apart), (b) that the ratio of propensity score variances between groups is close to 1.0, and (c) that the ratio of the residual variances of each covariate, after adjusting for the propensity score, must be close to 1.0. Residual variance is defined operationally as the original covariate regressed on the linear combination of the covariates that defines the estimated propensity score. The residual of this regression was examined for the variance ratios of these residuals between the two groups. Rubin (2001) suggested that a ratio of less than 0.5 or greater than 2.0 is considered extreme. A range of odds ratios between 0.8 and 1.2 represent equivalent variances.

(4) Cases were next ranked by their propensity scores, and stratified in quintiles with equal ranges (e.g., 0-0.2, 0.21-0.4, 0.41-0.6, 0.61-0.8, and 0.81-1.0). The assigned subclasses were coded as a new variable for each case. Sub-classes were examined further for a sufficient number of cases in each quintile. If either group has few cases in a quintile, that quintile will not be eligible for further evaluation.

[13] Although parsimony is often desirable, the technical recommendation from Rubin (1997) is to use a liberal number of variables to avoid any under-correction.

(5) Risks were tested for significant differences between groups by comparing the likelihood ratio χ^2 of a two-way log linear analysis of group by risk, with the likelihood ratio χ^2 of a three-way log linear analysis of group by risk when the model also includes the quintile plus all interactions (this is designated as "quintile plus" in the tables). The log linear model partitions categorical effects, similar to an analysis of variance. In other words, the model of cell frequencies explains why there are more scores in some cells than in others. A non-significant likelihood ratio χ^2 statistic in the three-way model for group by risk indicates that the addition of quintiles produces an optimal model that explains the pattern of cell frequencies. In other words, if the initial group difference is no longer significant when the quintiles and other group interactions are included, then the propensity score has adequately balanced the differences in risk profiles across groups.

If all covariates are balanced between FBCs and Hospital groups, then the sub-classification has successfully balanced pre-treatment covariates. However, if covariates are not successfully balanced in a subclass, common suggestions are that the logit model is modified to include higher order terms and the sub-classification procedures are re-evaluated for sufficient balance between groups (Dehejia & Wahba, 2002). This process is continued until all subclasses have balanced covariates between FBCs and Hospital groups.

The propensity score technique condensed the multiple biopsychosocial risks into one scalar function to produce an ignorable treatment assignment. And if the propensity score model was strong enough to balance all risks between groups, then theoretically all pre-existing observed differences no longer influence the treatment effects. The final adjustment model used the propensity score to control for group difference in the risk

profile overall and within each quintile. The following logistic regression model was used for calculating odds ratios and 95% confidence intervals for each outcome as a function of treatment group: Where $Y = b_0 + b_1X_1 + b_2X_2$; Y = Outcome; X_1 = group and X_2 = propensity score. In this study, interpreting whether the FBC service provides safe childbirth services depends on the extent to which poor infant and maternal childbirth outcomes are more likely to occur in the FBC group as compared to the Hospital comparison groups. Each outcome was evaluated on whether the odds ratio is close to 1.0. The 95% confidence interval allowed a straightforward assessment of whether each odds ratio, or the odds of occurrence in one group as compared to the other, was statistically significant. As discussed previously, being a member of the FBC group was coded as 1 and the incidence of having any risk or outcome was also coded as 1. Therefore, an odds ratio of less than 1.0 indicated that the FBC group was less likely to have an incidence of the risk or outcome than the Hospital group, while an odds ratio of greater than 1.0 indicated that the FBC group was more likely to have an incidence of the risk or outcome than the Hospital group. If the range of the confidence interval crossed 1.0 (i.e. 95%CI: 0.8-1.2), then the difference in likelihood of poor outcomes was not meaningful between groups.

CHAPTER IV

RESULTS

Results from this study provided evidence that freestanding birth center services are as safe as hospital-based services for low risk pregnant women, controlling for differences in risk factors. In support of this claim, the current study used two different comparison groups and several statistical techniques to control for differences on risk profiles between groups. Results are presented in two sections. The first section reports outcomes from comparing the FBC group to the low risk Hospital group (FBC-Low). The second section reports outcomes from comparing the FBC group to a midwife-attended Hospital group (FBC-Midwife).

For both FBC-Low and FBC-Midwife sections, results are reported in a similar sequence. First, differences between the two groups' risk profiles are reported in order to assess the extent to which selection bias is a possible problem in the interpretation of outcomes. Second, the difference between groups in the proportion of poor pregnancy outcomes are reported as univariate odds ratios as well as odds ratios adjusted for all risk factors discussed in the risk profile. For the FBC-Midwife analysis only, a more conservative statistical method using propensity scores was also employed to balance the overall risk profiles between groups. Propensity scores were calculated and their distributions were evaluated for sufficient overlap between FBC and midwife-attended Hospital groups. The propensity scores, or individual probabilities, were calculated by predicting FBC group membership by the risk profile. Therefore, examining the

propensity scores between groups reflects whether both groups have similar distributions on a scale that summarizes all observed risks. The propensity scores were then used to create five, equal, sub-categories in order to more closely compare outcomes within each category where propensities between groups were even more closely matched on observed risks. Outcomes adjusted for propensity scores and outcomes within each quintile are reported. A composite of these analyses helps guide interpretation of the overall set of morbidity and mortality outcomes between FBC and Hospital childbirth services, and reveals information about the stability of odds ratios on each of the outcomes.

Results: FBC versus Low Risk Hospital Group (FBC-Low)

Results comparing FBCs with low risk Hospital groups (FBC-Low) served as a benchmark for evaluating the safety of FBCs as compared to outcomes from previous research. The sampling frame for the FBC-Low analysis was similar to previous work comparing FBC outcomes with low risk pregnant women using Hospital services as discussed in Methods, Chapter 3. The low risk criteria were adapted from three studies that served as non-empirical comparison groups with outcomes from the National Birth Center Study (Rooks et al, 1989).

How Different are the Groups?

Examining risk profiles is an important first step in assessing whether the low risk Hospital group was an appropriate comparison for evaluating the safety of prenatal and childbirth services at FBCs. The overall pattern or risk revealed that the FBC group had

more demographic, behavioral, and medical risks than the Hospital group. The FBC-Low groups were statistically different on eight out of the 16 observed risks (Table 5).

Table 5: Risk Profile for FBC-Low Comparison

Risks	FBC, n = 2463		Hospital, n = 2500		Univariate Odds Ratio	95% confidence interval
	Frequency	Proportion	Frequency	Proportion		
<19 years old	237	0.096	163	0.033	1.53*	1.24-1.88
>35 years old	207	0.084	178	0.071	1.20	0.97-1.47
<12 years education^	569	0.231	505	0.202	1.19*	1.04-1.36
Race (African American)	93	0.038	253	0.101	0.35*	0.27-0.45
Unmarried^	634	0.257	603	0.241	1.09	0.96-1.24
> 20 weeks (5 months) gestation at first prenatal care visit+	125	0.051	127	0.052	0.99	0.77-1.27
Cigarette smoking	521	0.212	469	0.188	1.16*	1.01-1.34
Drinking alcohol^	16	0.007	26	0.010	0.62	0.33-1.16
Use of drugs++	31	0.013	37	0.016	0.84	0.52-1.35
First pregnancy	792	0.322	917	0.367	0.82*	0.73-0.92
Previous infant >4000 grams	73	0.030	48	0.019	1.56*	1.08-2.26
Previous small for gestation	13	0.005	8	0.003	1.66	0.69-4.00
Previous preterm	51	0.021	39	0.016	1.34	0.88-2.04
>2 previous terminations^	19	0.008	19	0.008	1.02	0.54-1.92
Any incidence of medical risk^	222	0.090	36	0.014	6.79*	4.75-9.70
Other category for medical risk^	116	0.047	383	0.153	0.27*	0.22-0.34

*p<0.05; ** p<0.01; *** p<0.001
^ Less than 6 cases total missing from either group
+ 4 cases missing from FBC group; 33 cases missing from Hospital group
++ 79 cases missing from FBC group; 113 cases missing from Hospital group

Of the eight risks that were statistically different between groups, five have *greater* odds of occurring in the FBC group than in the Hospital group. These risks included: Being less than 19 years old (FBC: 0.096 versus Hospital: 0.033 OR: 1.53,

95%CI: 1.24-1.88), having less than 12 years of education (FBC: 0.231 versus Hospital: 0.202; OR: 1.19, 95%CI: 1.04-1.36), smoking cigarettes during pregnancy (FBC: 0.212 versus Hospital: 0.188; OR: 1.16, 95%CI: 1.01-1.34), having a previous infant greater than 4000 grams at birth (FBC: 0.030 versus Hospital: 0.019; OR: 1.56, 95%CI: 1.08-2.26), and any incidence of medical risk[14] (FBC: 0.090 versus Hospital: 0.014; OR: 6.79, 95%CI: 4.75-9.70).

Risk markers that occur significantly *less* often in the FBC group than in the Hospital group included: Being of African American race (FBC: 0.038 versus Hospital: 0.101; OR: 0.35, 95%CI: 0.27-0.45), experiencing a first pregnancy (FBC: 0.322 versus Hospital: 0.367; OR: 0.82, 95%CI: 0.73-0.92), and an incidence of medical risk categorized in the general "other" category[15] (FBC: 0.047 versus Hospital: 0.153; OR: 0.27, 95%CI: 0.22 – 0.34). This risk profile pattern is contrary to the assumption that an FBC group has women with fewer risks than a group of low risk women in a Hospital group.

The two groups were not statistically different in the remaining eight risk factors. In other words, the 95% confidence interval around the odds ratio crosses 1.0. For example, the odds of being older than 35 years are 1.2 times greater in the FBC group than in the Hospital group. However, the 95% confidence interval (0.97-1.47) crosses over 1.0 indicating that this difference is not statistically meaningful and any difference

[14] Incidence of medical risk indicates whether any of the following medical risks are present for each case. Medical risks included in this variable are: anemia, cardiac disease, acute or chronic lung disease, diabetes, active genital herpes, hemoglobinopathy, chronic hypertension, renal disease, hydramnios or polyhydramnios, oligohydramnios, hypertension, pregnancy-induced, eclampsia, incompetent cervix, Rh sensitization, and uterine bleeding

[15] The risk factor labeled "other" represents other medical risk factors experienced by the mother that may cause or contribute to complications of this pregnancy. Examples are AIDS, preeclampsia, rubella, syphilis, gonorrhea, early onset of delivery and mental disorder.

would occur by chance alone. The remaining risks that are not likely to differentially occur in either group included: Drinking alcohol during pregnancy (FBC: 0.0065 versus Hospital: 0.0104; OR: 0.62. 95%CI: 0.33-1.16), having a previous small for gestational age infant (FBC: 0.0053 versus Hospital: 0.0033; OR: 1.66, 95%CI: 0.69-4.00), having greater than two previous terminations of a pregnancy (FBC: 0.0077 versus Hospital: 0.0076; OR: 1.02, 95%CI: 0.54-1.92), being unmarried (FBC: 0.26 versus Hospital: 0.24; OR: 1.09, 95%CI: 0.96-1.24), starting prenatal care late (FBC: 0.05 versus Hospital: 0.05; OR: 0.99, 95%CI: 0.77-1.27) using drugs during pregnancy (FBC: 0.01 versus Hospital: 0.02; OR: 0.84, 95%CI: 0.52-1.35), and having a previous preterm birth (FBC: 0.021 versus Hospital 0.016; OR: 1.34, 95%CI: 0.88-2.04)[16].

Missing cases within all the risk variables were minimal and did not significantly influence the outcomes. In both groups (FBC: n = 2463; Hospital: n = 2500), less than 0.3% of each risk is missing except for entering prenatal care late, which is missing in 33 cases (1.32%) of the Hospital group and drug use during pregnancy, which is missing in 133 cases (4.5%) of the Hospital group and 74 cases (3.2%) of the FBC group. There are 93 (3.78%) cases missing any risk in the FBC group and 150 cases (6.00%) in the Hospital group. The influence of missing cases was found to have minimal influence on each outcome. Therefore, these cases were dropped from the overall analysis.

Consistent with the earlier discussion of limitations in the 1989 National Birth Center Study (Rooks et al., 1989; Lieberman et al., 1989), the non-randomized treatment groups in this study are statistically different on many risk factors. The existence of

[16] It is important to point out that for many of the variables that are non-significant between groups, the proportion of occurrence found in either group is about 0.02 or less. Therefore, the inability to find a statistically significant difference may be the result of the lack of statistical power, given the very small proportions, than an accurate measure of group differences. For this reason, all risks will remain as part of the overall risk profile during statistical adjustments of outcomes.

differential risk profiles provide empirical support for the critique that selection bias may impede the interpretation of statistical estimates of effects between FBC and low-risk Hospital comparison groups. However, unlike previous critiques that suggest that FBC women are overall healthier than women in Hospital control groups, this analysis reveals the opposite trend. The FBC group has greater odds of having the majority of risks than the Hospital group, although many of the odds ratios are close to 1.0. However, the exclusion criteria for developing a low risk profile in the Hospital comparison group eliminates many of the medical risks, thereby producing a comparison where the FBC based women are actually "more medically risky" than those in the Hospital group on the aggregated medical risk factor (OR: 6.79, 95% CI: 4.75-9.70). The differences between groups also provide empirical support that control for these risks when comparing the FBC group with the Hospital control group, is a necessary statistical step in evaluating safety of FBC services.

Reported Outcomes Between FBC and Low Risk Hospital Groups

The majority of comparisons of mortality and morbidity rates between FBC and Hospital groups revealed that the probability of poor outcomes are less likely to occur in the FBC group than in the Hospital group (Table 6). As predicted, mortality rates are so rare in either group that neither neonatal death nor infant death was included in any further analyses in this study. Incident rates of neonatal death (1 case in FBC versus 7 cases in Hospital) and infant death (2 cases in FBC versus 5 cases in Hospital) are not statistically different between groups. The proportion of cases for seven out of the nine morbidity outcomes are also low, occurring in less than 0.10 of the cases in either FBC or

Hospital group. Outcomes with higher rates of occurrence included: Large for gestational age (0.127 in the FBC group versus 0.097 in the Hospital group), abnormal conditions of the infant (6.14% in the FBC group versus 13.17% in the Hospital group), and incidence of maternal complications during labor (17.89% in the FBC group versus 37.24% in the Hospital group).

Table 6: Outcomes, Frequency, Proportions and Univariate Odds Ratios with 95% Confidence Intervals for the FBC-Low Comparison

Outcomes	FBC, n = 2463		Hospital, n = 2500		Univariate Odds Ratio	95% Confidence Interval
	Frequency	Proportion	Frequency	Proportion		
Neonatal death	1	0.0004	7	0.003	0.15	0.02-1.18
Infant death	2	0.0008	5	0.002	0.41	0.08-2.09
LBW (<2500)	55	0.0223	149	0.060	0.36***	0.26-0.49
Small for gestational age^	157	0.0637	210	0.084	0.74**	0.60-0.92
Large for gestational age^	312	0.1267	241	0.097	1.36***	1.14-1.62
Preterm birth (<37)^	76	0.0309	209	0.084	0.35***	0.27-0.46
Apgar score, <7 at 1^	148	0.0602	202	0.081	0.73**	0.58-0.91
Apgar score, <7 at 5^	17	0.0069	23	0.009	0.75	0.40-1.41
Fetal distress^	63	0.0256	248	0.099	0.24***	0.18-0.32
Incidence of abnormal conditions of the infant^	151	0.0614	329	0.132	0.43***	0.35-0.53
Incidence of maternal complications during labor^	440	0.1789	930	0.372	0.37***	0.32-0.42

* p<0.05; ** p<0.01; *** p<0.001

^ Less than 5 cases missing from either FBC or Hospital group

Cases with missing outcomes were also dropped from further analysis after assessing their influence. Even fewer cases have missing outcome information than those with missing risk information[17].

All morbidity outcomes, except for large for gestational age (OR: 1.36, 95%CI: 1.14-1.62), have less odds of occurring in the FBC group than the Hospital group without any adjustment for unbalanced risk profiles. Women in the FBC group have significantly *less* odds of the following: Low birth weight (OR: 0.36, 95%CI: 0.26-0.49), small for gestational age (OR: 0.74. 95%CI: 0.60-0.92), preterm births (OR: 0.35, 95%CI: 0.27-0.46), Apgar scores less than seven at one minute (OR: 0.73. 95%CI: 0.58-0.91), fetal distress (OR: 0.24, 95%CI: 0.18-0.32), incidence of abnormal conditions of the infant (OR: 0.43, 95%CI: 0.35-0.53), and incidence of maternal complications (OR: 0.37, 95%CI: 0.32-0.42). An Apgar score less than seven at five minutes is the only outcome that is not significantly different between groups (OR: 0.75, 95%CI: 0.40-1.41).

Before the likelihood of poor pregnancy outcomes in the FBC group is fully interpreted, the difference in risk profiles between groups (Table 5) warrants additional analyses that control for the influence of risk on the outcomes. It is not possible to interpret whether the pattern of outcomes is driven by the differences in risks or the actual treatment effect without controlling for pregnancy risks. A direct, multivariate logistic model was used to predict the probability of each outcome as a function of using a FBC service after controlling for all 16 risk factors reported in the risk profile (Table 5).

[17] A proportion of less than 0.005 of the total cases (N = 4963) are missing outcome information.

Controlling for Risks Using Multivariate Logistic Regression

After taking into account the group differences in the risk profile, women in the FBC group remain less likely to experience all markers of morbidity than women in the Hospital group (Table 7). The odds ratio and 95% confidence intervals for each adjusted outcome do not markedly change the interpretation of the univariate odds ratios, although they all move closer to 1.0. Although the odds ratio point estimate for having a large for gestational age infant does not change dramatically, the likelihood of occurrence is no longer significantly different between the two groups (OR: 1.21, CI: 0.99-1.47). The likelihood of occurrence for the majority of outcomes remain significantly *lower* in the FBC group than in the Hospital group for outcomes such as: Low birth weight (OR: 0.39, CI: 0.27-0.55), small for gestational age (OR: 0.70, CI: 0.55-0.89), preterm birth (OR: 0.36, CI: 0.27-0.49), fetal distress (OR: 0.24, CI: 0.18-0.33), incidence of abnormal infant conditions (OR: 0.44, CI: 0.35-0.54), and incidence of maternal complication (OR: 0.38, CI: 0.33-0.44). Apgar scores that are less than seven at one minute are no longer statistically significant between groups (OR: 0.88, 95%CI: 0.67-1.09). Having an Apgar score less than seven at five minutes remains non-significant between groups, as in the univariate analysis; however, the point estimate also moves closer to 1.0 (OR: 0.87, 95%CI: 0.42-1.80).

Table 7 reports results of likelihood ratio tests for each outcome model. The log likelihood technique (-2 log likelihood) tests the overall *lack of model fit* by measuring the difference in the deviance (i.e. the amount of prediction not accounted for by the model) between the null model and the full model that includes all predictors (i.e. the influence of group and all risks) for each outcome (Cohen, Cohen, West, & Aiken, 2003).

For each morbidity outcome, the amount of deviance in prediction accounted for by the full model is less than the amount of deviance accounted for by the null model. For example, for low birth weight, the null –2 log likelihood model is equal to 1592.68. Whereas, when all predictors are included, the full –2 log likelihood model, the value equals 1446.55. This difference indicates that, collectively, the predictors contribute to the overall prediction of each outcome.

Table 7: Outcomes Adjusted for Risk Profile in FBC-Low Comparison

Outcomes[b]	-2 Log Likelihood		Likelihood Ratio Test				
	Intercept only model (null model deviance)	Intercept plus covariates model (full model deviance)	Chi-Square[a]	Pseudo R-Square	Max-rescaled R-Square	Adjusted Odds Ratio for group	95% Confidence Interval
LBW (<2500)	1592.68	1446.55	146.13**	0.03	0.11	0.39***	0.27-0.55
Small for gestational age	2473.06	2311.66	161.39**	0.03	0.08	0.70**	0.55-0.89
Large for gestational age	3314.11	3110.66	203.44**	0.04	0.08	1.21	0.99-1.47
Preterm birth (<37)	2045.63	1879.84	165.79**	0.03	0.10	0.36***	0.27-0.49
Apgar score, <7 at 1	2365.32	2326.82	38.51**	0.01	0.02	0.85	0.67-1.09
Apgar score, <7 at 5	422.71	390.58	32.13*	0.01	0.08	0.87	0.42-1.80
Fetal distress	2243.41	2058.69	184.71**	0.04	0.10	0.24***	0.18-0.33
Incidence of abnormal conditions of the infant	2987.29	2835.21	152.08**	0.03	0.07	0.44***	0.35-0.54
Incidence of maternal complications during labor	5492.91	5201.35	291.56**	0.06	0.09	0.38***	0.33-0.44

*p<0.05; ** p<0.01; *** p<0.001

[a] All with 17 degrees of freedom;

[b] The proportion of total cases dropped from each logistic regression model because of missing information is approximately 0.05.

The likelihood ratio test, however, also tests whether the overall prediction for each outcome model is greater than what you would expect by chance. Statistically significant values for $\chi 2(17, N = 4963)$, $p < 0.05$ for each outcome reveals that the amount of prediction is systematically larger than expected by chance alone (Cohen, Cohen, West, and Aiken, 2003). Both pseudo R-square and max-rescaled R-square[18] indices, or the percent of null deviance in the outcome that is accounted for by the full model, are small. The max-rescaled R-square statistic indicates that the model reliably accounts for only 11% of the deviance in predicting low birth weight and 10% of the deviance in predicting preterm birth. For all other outcomes, less than 10% of the null deviance is accounted for by the full set of predictors (group plus all risks).

The FBC-Low comparison used similar criteria as previous studies to construct Hospital comparison groups and found similar results as documented in previous studies. Despite differences in risk profiles revealing that the FBC group had greater risks on the majority of demographic, behavioral, and medical factors, after adjusting for these differences poor outcomes are more likely to occur in the Hospital group. However, differences between groups such as provider practices and individual attitudes or beliefs are unobserved in this analysis, and could further confound the reported treatment differences. Therefore, incorporating procedural (midwife practice models) and dispositional (personal attributes associated with choosing birth center services) commonalities between FBC and Hospital comparison groups have the potential to further minimize bias in treatment differences. The second analysis, therefore, compares

---

[18]Because the pseudo R squared achieves a maximum level of less than 1, for discrete models, the max rescaled R-square (Nagelkerke, 1991) was proposed as an adjusted coefficient that can reach a maximum value of 1.Where $R^2 = 1 - \{L(o)/L(B)\}^{2/n}$ and $R_{max}^2 = 1 - \{L(0)\}^{[2/n]}$, the max rescaled R-squared equals $1-(R^2 / R_{max}^2)$ from Nagelkerke (1991).

the FBC group to a midwife-attended Hospital group in order to better control for these commonalities.

Results: FBC versus Midwife-attended Hospital Services (FBC-Midwife)

The sampling frame for comparing FBC services to midwife-attended Hospital services controls for provider practice differences documented in the current research. In addition, limiting the control group to women with midwife-attended hospital births also controls for unobserved physical or psychological risk factors influential in the choice of midwife versus physician providers. Results are reported in the same sequence as for the FBC-Low comparison.

How Different are the Groups?

The risk profile comparing FBC and midwife-attended Hospital groups (Table 8) reveals that there are more differential risks between groups than when comparing FBC and low risk Hospital groups. Although the groups may be more closely matched on unobserved factors connected with provider practice and characteristics related to choice of provider, 11 out of the 16 risks factors are significantly different between the FBC and midwife-attended Hospital groups. Also, the pattern of risks is different between the two analyses. In FBC-Low comparison there is more overall risks occurring in the FBC group than in the Hospital group; the FBC-Midwife analysis presents an opposite pattern of risk. Of the 11 significantly different risks, eight occur more often in the Hospital group than in the FBC group. The FBC women have *lower* odds of being younger than 19 years (FBC: 0.096 versus Hospital: 0.121; OR: 0.77, 95%CI: 0.67-0.89), having less

than 12 years of education (FBC: 0.231 versus Hospital: 0.269, OR: 0.82, 95%CI: 1.42-1.97), being of African American race (FBC: 0.038 versus Hospital: 0.099; OR: 0.36, 95%CI: 0.29-0.44), being unmarried (FBC: 0.257 versus Hospital: 0.305; OR: 0.79, 95%CI: 0.71-0.87), starting prenatal care late (FBC: 0.051 versus Hospital: 0.067; OR: 0.75, 95%CI: 0.62-0.91), having a first pregnancy (FBC: 0.322 versus Hospital: 0.372; OR: 0.8, 95%CI: 0.73-0.88), having an incidence of medical risk (FBC: 0.090 versus Hospital: 0.164; OR: 0.51, 95%CI: 0.44-0.59) and having any incidence of medical risk in the general "other" category (FBC: 0.047 versus Hospital: 0.100; OR: 0.45, 95%CI: 0.37-0.55). Risks that appear significantly *more* often with FBC women include being older than 35 years (FBC: 0.084 versus Hospital: 0.052;OR: 1.68. 95%CI: 1.42-1.97), having a previous birth with infant weight greater than 4000 grams (FBC: 0.030 versus Hospital: 0.015; OR: 1.97, 95%CI: 1.49-2.60), and having a previous preterm infant (FBC: 0.021 versus Hospital: 0.013; OR: 1.56, 95%CI: 1.13-2.15). The risk profile for FBC-Midwife provides empirical support for the methodological concern that an FBC group is an overall "healthier" group of women than those in the Hospital control group.

Five risks are not statistically different between the FBC and midwife-attended Hospital groups. The two groups have balanced proportions on all three behavioral risk factors including cigarette smoking (FBC: 0.212 versus Hospital: 0.196; OR: 1.09, 95%CI: 0.99-1.22), alcohol use (FBC: 0.007 versus Hospital: 0.007; OR: 0.92 95%CI: 0.54-1.58), and drug use during pregnancy (FBC: 0.013 versus Hospital: 0.010; OR: 1.25 95%CI: 0.84-1.86). There are also an equivalent number of women with previously small for gestational age infants (FBC: 0.005 versus Hospital: 0.003; OR: 1.58, 95%CI: 0.84-2.97) and women who have had greater than two previous pregnancies terminated

(FBC: 0.008 versus Hospital: 0.005; OR: 1.67, 95%CI: 0.99-2.82)[19]. Despite five of the risks being equivalently distributed across both groups, there are enough differences to also warrant the need for additional statistical adjustments in order to minimize bias when interpreting outcomes.

Table 8: Risk Profile for FBC-Midwife Comparison

Risks markers	FBC, n = 2463		Hospital, n = 11,635		Univariate Odds Ratio	95% CI
	Frequency	Proportion	Frequency	Proportion		
<19 years old	237	0.096	1407	0.121	0.77***	0.67-0.89
>35 years old	207	0.084	604	0.052	1.68***	1.42-1.97
<12 years education^	569	0.231	3132	0.269	0.82***	0.74-0.90
Race (African American)^	93	0.038	1149	0.099	0.36***	0.29-0.44
Unmarried	634	0.257	3552	0.305	0.79***	0.71-0.87
> 20 weeks (5 months) gestation at first prenatal care visit+	125	0.051	772	0.067	0.75**	0.62-0.91
Cigarette smoking^	521	0.212	2281	0.196	1.09	0.99-1.22
Drinking alcohol^	16	0.007	82	0.007	0.92	0.54-1.58
Use of drugs++	31	0.013	120	0.010	1.25	0.84-1.86
First pregnancy^	792	0.322	4323	0.372	0.8*	0.73-0.88
Previous infant >4000 grams^	73	0.030	178	0.015	1.97***	1.49-2.60
Previous small for gestation^	13	0.005	39	0.003	1.58	0.84-2.97
Previous preterm^	51	0.021	156	0.013	1.56**	1.13-2.15
>2 previous terminations^	19	0.008	54	0.005	1.67	0.99-2.82
Any incidence of medical risk^	222	0.090	1910	0.164	0.51***	0.44-0.59
Other category for medical risk^	116	0.047	1157	0.100	0.45***	0.37-0.55

*p<0.05; ** p<0.01; *** p<0.001
^ Less than 6 cases total missing from either group
+ 4 cases missing from FBC group; 30 cases missing from Hospital group
++ 79 cases missing from FBC group; 122 cases missing from Hospital group

[19] Note that, once again, even though the odds ratios point estimates are rather large, the proportion of risk in either group for alcohol use, having a previously small for gestation age infant, and having had greater than two previous pregnancies is extremely rare (less than 1%). Therefore, the inability to find a statistically significant difference in these risks may be attributed to low statistical power.

Reported Outcomes Between FBC and Midwife-Attended Hospital Groups

The morbidity and mortality outcomes in FBC-Midwife show a very similar pattern to FBC-Low, despite the differences in risk profiles between the two analyses (Table 9). Once again infant mortality rates for the Hospital group occur very rarely and will not be adjusted for by any additional statistical model.

Table 9: Outcomes, Frequencies, Proportions, and Univariate Odds Ratios with 95% Confidence Intervals for FBC-Midwife Comparison

Outcomes	FBC, n = 2463		Hospital n = 11,635		Univariate Odds Ratio	95% CI
	Frequency	Proportion	Frequency	Proportion		
Neonatal death	1	0.0004	10	0.0009	0.47	0.06-3.69
Infant death	2	0.0008	16	0.0014	0.59	0.14-2.57
LBW (<2500)	55	0.022	388	0.033	0.66**	0.50-0.88
Small for gestational age	157	0.064	929	0.080	0.79**	0.66-0.94
Large for gestational age	312	0.127	968	0.083	1.60***	1.40-1.83
Preterm birth (<37)^	76	0.031	609	0.052	0.58***	0.45-0.74
Apgar score, <7 at 1 minute^	148	0.060	694	0.060	1.01	0.84-1.21
Apgar score, <7 at 5 minutes^	17	0.007	76	0.007	1.06	0.62-1.79
Fetal distress^	63	0.026	980	0.084	0.29***	0.22-0.37
Incidence of abnormal conditions of the infant^	151	0.061	829	0.071	0.85	0.71-1.02
Incidence of maternal complications during labor^	440	0.179	2461	0.212	0.81***	0.73-0.91

*p <0.05; **p<0.01; ***p<0.001
^ Less than 5 cases missing from either FBC or Hospital group

Without controlling for risk profiles, six of 11 morbidity outcomes are significantly different between groups. The exceptions are neonatal and infant death, Apgar scores of less than seven at either one minute or five minutes, and the incidence of

abnormal infant conditions. Only having an infant that is large for gestational age occurs significantly more often in the FBC group (0.127) than the Hospital group (0.083), with an odds ratio of 1.60 (95%CI: 1.40-1.83)[20]. Incidence of low birth weight (FBC: 0.022 versus Hospital: 0.033; OR: 0.66, 95%CI: 0.50-0.88), small for gestational age (FBC: 0.064 versus Hospital: 0.080; OR: 0.79, 95%CI: 0.66-0.94), preterm birth (FBC: 0.031 versus Hospital: 0.052; OR: 0.58, 95%CI: 0.45-0.74), fetal distress (FBC: 0.026 versus Hospital: 0.084; OR: 0.29, 95%CI: 0.22-0.37), incidence of abnormal conditions (FBC: 0.061 versus Hospital: 0.071; OR: 0.85, 95%CI: 0.71-1.02), and maternal complications (FBC: 0.179 versus Hospital: 0.212; OR: 0.81, 95%CI: 0.73-0.91) all occur significantly more often in the Hospital group than the FBC group. Similar to FBC-Low, missing cases were minimal for both risks and outcomes and therefore no imputation methods were performed[21].

Controlling for Risks Using Multivariate Logistic Regression

Controlling for all 16 risks by using a direct logistic regression model also reveal similar results as in the FBC-Low analysis (Table 10). As expected, by comparing FBC to midwife-attended hospital births, the majority of odds ratios for each outcome move closer to 1.0. The trends and interpretations for low birth weight, small for gestational age, preterm birth, fetal distress, and incidence of maternal complications during labor do not change from what is seen in the FBC-Low comparison. These outcomes are all significantly more likely to occur in the Hospital group than in the FBC group. The

[20] Of the 312 women with large for gestation age infants, 29 (proportion = 0.093) cases where transfer cases and delivered in the hospital

[21] Missing information made up a proportion of the total sample of less than 0.015 for any risk variable and 0.0005 for any outcome variable.

results of having an Apgar score less than seven at either one or five minutes also remain non-significant between the FBC-Midwife groups, which is consistent with the results found in the FBC-Low comparison. Given that the odds ratios move closer to being statistically equivalent (OR: 1.0) for all of the above poor pregnancy outcomes, it is possible that matching groups on unobserved characteristics related with provider characteristics further minimizes bias in the overall estimates of treatment effect.

Table 10: Outcomes Adjusted for Risk Profile in FBC-Midwife Comparison

Outcomes[b]	-2 Log Likelihood		Likelihood Ratio Test				
	Intercept only model (null model deviance)	Intercept plus covariates model (full model deviance)	Chi-Square[a]	Pseudo R-Square	Max-rescaled R-Square	Adjusted Odds Ratio for group	95% Confidence Interval
LBW (<2500)	3900.644	3619.235	281.409***	0.020	0.082	0.72*	0.54-0.98
Small for gestational age	7513.561	7004.288	509.274***	0.036	0.086	0.80*	0.66-0.96
Large for gestational age	8394.779	7903.466	491.313***	0.035	0.077	1.47***	1.27-1.69
Preterm birth (<37)	5429.838	5180.613	249.225***	0.018	0.055	0.59***	0.46-0.76
Apgar score, <7 at 1	6302.007	6224.495	77.512***	0.006	0.015	1.07	0.89-1.30
Apgar score, <7 at 5	1105.85	1078.813	27.037	0.002	0.025	1.13	0.65-1.96
Fetal distress	7367.873	7018.477	349.396***	0.025	0.060	0.31***	0.24-0.40
Incidence of abnormal conditions of the infant	7016.006	6664.617	351.390***	0.025	0.063	0.99	0.82-1.20
Incidence of maternal complications during labor	14079.03	13728.76	350.269***	0.025	0.039	0.84**	0.75-0.95

*p<0.05; ** p<0.01; *** p<0.001

[a] All with 17 degrees of freedom

[b] The proportion of total cases dropped from each logistic regression model because of missing information is approximately 0.019.

The only two outcomes that differ between the two comparison groups are having an incidence of abnormal conditions for the infant and being large for gestational age. In the FBC-Midwife comparison, the incidence of infant abnormal conditions is no longer significantly different between groups (OR: 0.99, 95% 0.82-1.20). Large for gestational age, however, shows the opposite trend where in the FBC-Midwife comparison there is now a significant difference between groups. The FBC group has significantly greater odds of having a large for gestational age baby than in the midwife-attended Hospital group (OR: 1.47, 95%CI: 1.27-1.69)[22].

An examination of model deviance tests (-2 log likelihood) finds that the set of predictors for the majority of outcomes collectively contribute to the overall prediction of each outcome. The amount of deviance in prediction accounted for by the full model is less than the amount of deviance accounted for by the null model. And, prediction models are larger than expected by chance alone $\chi 2(17, N = 4963)$, $p < 0.05$ for every outcome except having a low Apgar score at five minutes. For a low Apgar score at five minutes, the contribution of the model for predicting deviance is not larger than chance ($\chi 2(17, N = 4963) = 27.04$). The max-rescaled R-square indicates that the predictors account less than 9% of the deviance in any model. The remaining deviance to be predicted for most outcomes, in addition to the possibility that logistic regression techniques are not trustworthy due to the large number of predictors, provides a rationale for employing a secondary statistical adjustment using the propensity score technique.

[22] The implications for FBCs having greater proportions of larger for gestational age (LGA) infants are discussed in detail in the discussion section. As described, it is not clear whether LGA is a negative outcome because it is not related to any additional complication.

Developing a Propensity Score Model

As discussed in Methods, Chapter 3, Rubin (2001) suggested that predicting outcomes between nonrandomized groups could be problematic when there are many covariates involved. Researchers have found that if there are less than 10 outcomes per covariate, logistic regression models have an increase in the effect of bias (Peduzzi, Concato, Kemper, Holford, & Feinstein, 1996). In this situation, propensity scores have been found to produce estimates that are less biased than seen in logistic regression estimates (Cepeda et al., 2003). The problem develops from the difficulty in evaluating the often multi-dimensional effect of many small differences. Therefore, Rubin and others suggest that calculating propensity scores and evaluating their distributions between groups produce a more precise understanding of whether groups are sufficiently similar (i.e. overlap) on the full set of covariates (Dehejia & Wahba, 2002; Winship & Morgan, 1999; Rosenbaum & Rubin, 1984; Rubin, 1997).

Recall that propensity scores are the individual probabilities for each case of being a member of the FBC group as a function of the complete set of risk factors. Therefore, examining the distribution of propensity scores between groups provides a clear picture of how well the two groups overlap on the risk profile as a scalar function of the entire set of risk factors. For the FBC-Midwife sample, two propensity score models were needed. The first used only the original 16 risk factors as predictors in the propensity score model. After checking the distributions of overall propensity scores and within quintiles, complete balance of the risk profile was not achieved. Therefore, a second model was calculated using the original 16 risk factors and an additional 53 (Appendix B, Table B-2) eligible interaction variables to assess the outcomes with a more

precise estimate of treatment effects. The expanded model produces enough balance on risk profiles to warrant using the propensity scores as a second method for statistically controlling for pre-existing group differences when calculating poor pregnancy outcomes.

First Round: Simple Propensity Score Model.

The results from the first propensity model illustrates that the distributions of propensity scores between the FBC-Midwife comparisons appear to have considerable overlap between groups (Figure 1). However, only two of the three criteria proposed by

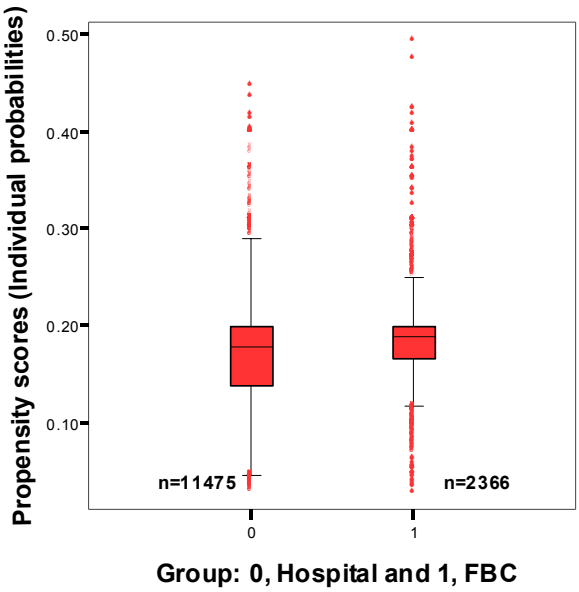

Figure 1: Distribution of Propensity Scores in the FBC-Midwife Comparison Using the First Propensity Score Model

Rubin (2001) are satisfied (Table 11). The ratio of the propensity score variance between groups is close to 1 (ratio of variance = 1.01) satisfying the first criterion. And the mean difference in the propensity score for FBC-Midwife (MeanFBC – MeanHospital =

0.0196) is less than half of the pooled standard deviation (SD_{pooled}= 0.026), satisfying criterion two. The ratios of the residual variances of the covariates after adjusting for the propensity score, however, are not all close to 1.0 (Appendix B, Table B-1). In fact, ratios for only 5 (0.31) of the 16 risks are considered equivalent ratios using the acceptable range (OR: 0.8-1.2) specified by Rubin (2001).

Table 11: Distribution of Propensity Scores with Odds Ratios for Residual Variances in First Propensity Score Model

Group	**N Obs**	**N**	**Mean**	**Std Dev**	**Variance**
0, Hospital	11635	11475	0.168	0.0518469	0.00269
1, FBC	2463	2366	0.187	0.0515372	0.00266
Range of odds ratios of residual variances	**</= 0.5**	**>0.5 - </= 0.8**	**>0.8 - </= 1.2**	**>1.2 - </= 2.0**	**>2.0**
# of variables	0	6	5	4	1
Proportion	0.00	0.38	0.31	0.25	0.06

After sub-classifying the propensity scores into quintiles there are no "empty" quintile cells for either FBC or Hospital groups that might signify that the overall group differences are too great for propensity score adjustment to solve. However, examination of the addition of quintiles also reveals that the propensity score technique comes close to balancing the FBC-Midwife comparison on the full risk profile, but does not successfully balance all risks. Results from comparing a two-way and three-way log linear model show succinctly the effect of propensity scoring across the two groups (Table 12). The two-way model, using only group and each risk variable, shows that 11 of the 16 risks have significant likelihood ratio χ^2 as previously established during the examination of risk profiles. The inclusion of quintiles in the three-way model eliminates all significant

confounding risks, except having less than 12 years of education, χ^2 (1, N =13,841) = 5.65, p <0.02. The unbalanced risk factor between groups provides additional evidence that the propensity score did not control for all pre-existing group differences.

Table 12: Results from the Log-Linear Analysis Comparing the Balance of Risks Between Groups With and Without Quintiles Using the First Propensity Score Model

Group*Risk	2-way log linear model Group*Risk		3-way log linear model Group*Risk*Quintile	
	Chi-Square	Pr>ChiSq	Chi-Square	Pr>ChiSq
<19 years old	11.99	0.0005	2.06	0.151
>35 years old	37.96	<.0001	0.02	0.895
<12 years education	15.03	0.0001	5.65*	0.018
African American risk marker	3219.14	<.0001	2.48	0.115
Unmarried	22.25	<.0001	1.71	0.190
>20 weeks (or 5 months) gestation at first prenatal care visit	8.31	0.0039	1.03	0.310
Cigarette smoking	3.04	0.0813	3.43	0.064
Drinking alcohol	0.09	0.7652	0.61	0.434
Use of drugs	1.22	0.2696	0.14	0.704
First pregnancy - parity risk	21.97	<.0001	0.15	0.700
Prev infant 4000+ grams	23.19	<.0001	0.02	0.892
Prev small for gestation	2.03	0.1538	0	0.960
Prev preterm	7.43	0.0064	0	0.997
Greater than 3 previous terminations - (miscarriage, abortion, still birth)	3.67	0.0555	0.48	0.487
Sum of medical risks	83.48	<.0001	0.11	0.739
Other	64.27	<.0001	0.31	0.576

*p <0.05

The fact that the residual ratios between groups and within quintiles remain unbalanced provides evidence that the propensity score model is not strong enough to correct for pre-existing differences in the risk profile. Therefore, a second propensity

score model was calculated using the full risk profile plus a set of interactions of eligible risks.

Second Round: The Expanded Propensity Score Model

Interactions were chosen for the second, expanded propensity score model by examining every interaction between all 16 risk factors within groups. No specific theory suggests which interactions are more meaningful for this set of risks; therefore, a scheme was created to systematically include eligible interactions in the expanded propensity model. If there was a significant relationship, using a 2x2 chi-square test, between two risks in either FBC or midwife-attended Hospital group, that interaction was included in the model that calculated the second set of propensity scores. There are 53 interactions eligible to include in the second propensity score model (Appendix B, Table B-2). By adding the interactions with the original risk factors, a total of 69 covariates are used to calculate the propensity scores using a direct logistic regression model.

Rubin's (2001) three criteria were used again to evaluate the trustworthiness of the larger set of covariates to successfully balance the FBC and Hospital group. Once again, the distributions of propensity scores between the two groups are quite similar (Figure 2). The extended propensity model successfully meets two of the three criteria. The ratio of propensity score variances between the FBC and Hospital groups is 0.80, and therefore is close to 1.0. The difference between propensity score means (MeanFBC – MeanHos = 0.035) is approximately equal to half of the pooled standard deviation (SD_{pooled} = 0.034), therefore marginally satisfying criterion two. And, the level of success for meeting the third criterion, that all ratios of the propensity score residuals

between groups is close to 1.0, improves for the larger propensity score model as compared to the original model (Appendix B, Table B-3). The majority of ratios (79%) are considered not extreme, i.e. falling between 0.5 and 2.0, and only 29 (42%) of the 69 total variables are considered equivalent (Table 13). If only the original 16 variables are examined, 9 (0.56) ratios are considered equivalent. However, the proportion of residual ratios that fall within an equivalent range with the expanded propensity model is greater than seen in the original propensity model where only 5 (0.31) variables had residual ratios that were equivalent between groups. Therefore, the larger propensity model is considered an improvement.

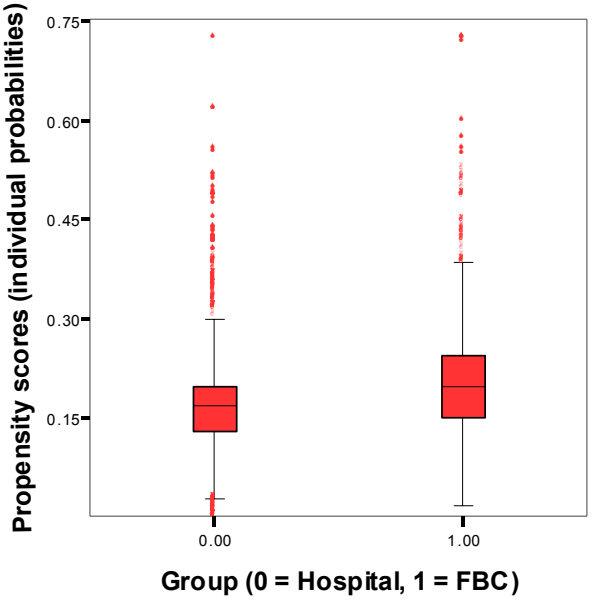

Figure 2: Distribution of Propensity Score in the FBC-Low Comparison Using the Extended Propensity Score Model

Table 13: Distribution of Propensity Scores with Odds Ratios for Residual Variances in the Extended Propensity Score Model

Group	**N Obs**	**N**	**Mean**	**Std Dev**	**variance**
0, Hospital	11635	11475	0.165	0.067	0.005
1, FBC	2463	2366	0.200	0.075	0.006
Range of odds ratios	**</= 0.5**	**>0.5 - </= 0.8**	**>0.8 - </= 1.2**	**>1.2 - </= 2.0**	**>2.0**
# of variables	2	8	29	17	13
Proportion	0.03	0.12	0.42	0.25	0.19

Although the three criteria are not completely satisfied with the addition of the interactions, the overall propensity model more successfully balances the FBC and Hospital groups across all five quintiles (Appendix B, Table B-4). The results from the two log linear analysis comparing the relationship of group by risk with a three way model that includes the quintile variable were examined closely. In all but two of the 69 risks and interactions, the relationship between group and each risk is not significantly different, $\chi 2(1, N = 14{,}098)$, $p < 0.05$ (Table 14). The interaction of having less than 12 years of education and cigarette smoking (edurisk*cigrisk, $p<0.02$) and the interaction of having a first pregnancy and cigarette smoking (parrisk2*cigrisk, $p<0.05$) are the only two variables that remain unbalanced between groups when quintiles are included in the analysis. However, it is important to note that neither interaction was significant in the 2-way model before the propensity distribution was divided into quintiles.

The expanded propensity score model does a better job balancing the risk profile between groups than the original propensity score model, even though over half of the risks and their eligible interactions have ratios of residual variance between groups that remain too large according to Rubin's (2001) criteria. Although there is not perfect balance between groups on the entire risk profile, the variation of group is only

marginally related to the propensity score. An examination of the overall correlation between group and propensity score (r = 0.188) reveals that it is not likely that any additional balancing of the risk profiles will produce widely different outcomes. The correlations for group and propensity scores within each quintile are also all quite small (Quintile 1: r = 0.095; Quintile 2: r = 0.027; Quintile 3, r = 0.006; Quintile 4, r = -0.025; Quintile 5, r = 0.147). Therefore, any additional statistical techniques employed to fully balance each risk within each quintile is not likely to alter the odds ratio for any outcome.

Table 14: Summary of Results from the Log-Linear Analysis Comparing the Balance of Risks Between Groups With and Without Quintiles Using the Extended Propensity Score Model (Appendix B, Table B-4)

Unbalance between groups	2 way log linear model		3 way log linear model
Group*Risks			
Yes, p < 0.05	37	p > 0.05	37
		Remain p<0.05	0
No, p >0.05	30		^2
n/a	2		

^2 risks now significantly different between groups in 3-way model.

Controlling for Risk Profiles Using Propensity Scores

Despite evidence that the propensity score significantly decreased the pre-existing differences between groups, controlling for risk profiles by using propensity scores does not markedly alter the interpretation of any outcome (Table 15). The likelihood of poor pregnancy outcomes such as low birth weight, small for gestational age, preterm birth, fetal distress, or the set of maternal complications occurring in the FBC group is lower than the likelihood of these outcomes occurring in the Hospital group. The likelihood of having low Apgar scores or having infant complications remains statistically equivalent between the two groups. And the likelihood of having a large for gestational age infant is 1.45 times greater in the FBC than in the Hospital groups.

Table 15: Summary of Odds Ratios for Each Outcome with All Adjustment Models in the FBC-Midwife Comparison

	LBW (<2500)	Small for gestational age	Large for gestational age	Preterm birth (<37)
Univariate	0.66* (0.50-0.88)	0.79* (0.66-0.94)	1.60* (1.40-1.83)	0.58* (0.45-0.74)
Adjusted for individual risks	0.72* (0.54-0.98)	0.80* (0.66-0.96)	1.47* (1.27-1.69)	0.59* (0.46-0.76)
Adjusted for propensity score extended model	0.74* (0.55-0.99)	0.80* (0.67-0.96)	1.45* (1.26-1.67)	0.61* (0.48-0.79)
Quintile 1 FBC, n = 211 Hosp, n = 2557	0.90(0.45-1.80)	0.61(0.34-1.09)	1.69(1.04-2.73)	0.63(0.34-1.18)
Quintile 2 FBC, n = 416 Hosp, n = 2566	0.78(0.40-1.51)	0.78(0.50-1.21)	1.44(1.02-2.04)	0.74(0.42-1.30)
Quintile 3 FBC, n = 395 Hosp, n = 1856	0.96(0.50-1.85)	0.87(0.59-1.29)	1.06(0.70-1.63)	0.47(0.22-0.97)
Quintile 4 FBC, n = 600 Hosp, n = 2463	0.67(0.26-1.72)	1.17(0.73-1.85)	1.40(1.10-1.78)	0.40(0.19-0.82)
Quintile 5 FBC, n = 744 Hosp, n = 2033	0.58(0.34-0.97)	0.71(0.52-0.98)	1.70(1.30-2.21)	0.70(0.46-1.05)

* p <0.05

Table 15 continued

	Apgar score, <7 at 1	Apgar score, <7 at 5	Fetal distress	Abnormal conditions	Maternal complications
Univariate	1.01 (0.84-1.21)	1.06 (0.62-1.79)	0.29* (0.22-0.37)	0.85 (0.71-1.02)	0.81*(0.73-0.91)
Adjusted for individual risks	1.07 (0.89-1.30)	1.13 (0.65-1.96)	0.31* (0.24-0.40)	0.99 (0.82-1.20)	0.84* (0.75-0.95)
Adjusted for propensity score extended model	1.07 (0.89-1.29)	1.08 (0.62-1.87)	0.31* (0.24-0.41)	1.01 (0.84-1.22)	0.85* (0.76-0.96)
Quintile 1	1.04(0.62-1.75)	1.81(0.62-5.29)	1.10(0.69-1.75)	0.67(0.42-1.09)	0.98(0.72-1.34)
Quintile 2	0.91(0.59-1.40)	0.81(0.18-3.54)	0.28(0.15-0.50)	1.45(0.98-2.17)	0.99(0.76-1.29)
Quintile 3	1.12(0.70-1.79)	1.60(0.43-5.97)	0.24(0.10-0.54)	0.89(0.53-1.50)	0.78(0.58-1.05)
Quintile 4	1.24(0.85-1.81)	1.06(0.30-3.80)	0.09(0.04-0.20)	1.09(0.72-1.63)	0.73(0.56-0.95)
Quintile 5	1.04(0.72-1.52)	0.73(0.24-2.19)	0.36(0.21-0.61)	1.02(0.70-1.47)	0.83(0.66-1.03)

* p <0.05

Interpretations of the lack of model fit tests also do not change between controlling for the risk profile model and the extended propensity score model (Appendix B, Table B-5).

Table 15 also reveals that the patterns of odds ratios stay consistent throughout all quintiles. Cases within each quintile are examined separately using the same logistic regression model to calculate odds ratios to investigate whether the treatment effect for any outcome is overtly different within the quintiles as compared to the full distribution of propensity scores. The decrease in statistical power when examining the individual quintiles makes a statistical test for significance problematic. Therefore, the confidence intervals of odds ratios are descriptively examined for overlap between the full distribution and each quintile. Examining the outcomes within each quintile helps evaluate whether cases in one end of the distribution of propensity scores have different odds ratios than the overall sample. Since the purpose of this study is to evaluate the safety of FBCs, it is important to know whether the pattern of odds ratios for each outcome stay consistent for all cases within the entire distribution of propensity score.

There are two outcomes, however, that show inconsistent findings in one or more quintiles. The occurrence of fetal distress is consistently and significantly more likely to occur in the Hospital group than the FBC group. However, the odds ratio falls outside the confidence interval range for the full distribution in quintile 1 (OR: 1.10, 95%CI: 0.69-1.75) and in quintile 4 (OR: 0.09, 95%CI: 0.04-0.20). Neither outcome is problematic for the FBC group. Upon further investigation there are 22 cases (0.104) with in quintile 1 as compared to the Hospital group with 245 cases (0.096). Whereas, in quintile 4 there are 6 cases in the FBC group (0.010) compared with 247 cases in the Hospital group (0.100). The odds ratio for abnormal conditions of the infant also falls

outside the range for the full distribution in quintile 2 (OR: 1.45; 95%CI: 0.98-2.17). The frequency of infant complications in the FBC group is 32 cases (0.08) as compared to the Hospital group with 140 cases (0.06).

Overall, the propensity score analysis does not alter the interpretations for outcomes as compared to the logistic model controlling for individual risk variables. However, the rigorous attention to attaining balance between groups provides additional strength to the overall interpretation of safety between FBC and Hospital services. And, the majority of comparisons within quintiles have consistent odds ratios between FBC and Hospital services on all outcomes. The examination of the quintiles provides a very detailed descriptive assessment of whether there are any troubling odds ratios throughout the propensity score distribution that might call into question the interpretation of safety. Using the theory of propensity scores (Rubin, 1997, 2001), it is reasonable to believe that the extent to which groups are balanced has minimized the bias attributed to the pre-existing risks measured in this study.

CHAPTER V

DISCUSSION

The appropriate location for childbirth has been debated for over 40 years between advocates for home-like environments and hospital environments. The question of whether childbirth services outside of the hospital can be a viable option in the maternal health care system hinges on whether services, such as freestanding birth centers (FBC), are safe for women and infants. Although previous studies have consistently reported low rates of poor childbirth outcomes in FBCs (Cordray & Davidson, 2004; Jackson et al., 2003; Rooks et al., 1989), methodological limitations common to any observational study have impeded the use of their evidence. In particular, the presence of selection bias has challenged interpretation of findings which compare FBC and Hospital outcomes. Selection bias is relevant because there are differential dispositional characteristics that drive women to choose the care of midwifery professionals and location for childbirth. The common assumption that women using FBC services are thought to be overall healthier than women designated as low risk using hospital services (Albers & Katz, 1991; Fullerton & Wingard, 1990; Jackson, Lang, Dickinson, & Fullerton, 1994; Lieberman & Ryan, 1989; Howell-White, 1997; Walsh & Downe, 2004).

The difficulty in conducting credible research exists in how best to design comparative hospital groups, which appropriately match medically low risk women who choose FBC services. On the surface, this observational design study is challenged by

the same generic problems of self-selection common in all non-randomized, comparison studies. However, the design of this study attempts to overcome the selection bias challenge by creating a second hospital comparison group of midwife-attended women in addition to the more common comparison group of medically low risk women. Comparing midwife-attended hospital births helps to control for the unobserved dispositional characteristics connected with choosing midwifery care. Furthermore, this study, with propensity score analyses, attempts to mimic the randomization process by balancing the observed pre-existing risk factors between groups. Both techniques help minimize bias attributed to differing risk profiles.

Overview of Findings from FBC-Low and FBC-Midwife Comparisons

The first analysis compared the FBC group to a Hospital group of low risk women. Assessment of risk profiles found that the two groups were different; therefore, providing empirical support to the concern that selection bias is a problem. Unexpectedly, the women in the FBC group had a *greater* proportion of demographic, behavioral and medical risks than the women in the Hospital group. However, after controlling for these higher risks, the FBC group had overall better outcomes than the Hospital group. Low birth weight, small for gestational age, preterm birth, fetal distress, incidence of abnormal conditions for the infant, and incidence of maternal complications were all significantly more likely to occur in the Hospital group than in the FBC group. The FBC group was just as likely as the Hospital group to have cases of infants with large birth weight for their gestational age and infants with Apgar scores less than seven at one and five minutes.

The comparison group of low risk women receiving hospital-based services was similar to the comparison groups found in previous studies. Even though the overall outcomes were consistent with previous studies, selection bias continues to be a potential problem because there is no control for dispositional characteristics associated with those women who chose midwifery care in and out of the hospital. This limitation provided the rationale for constructing the FBC-Midwife comparison.

The FBC-Midwife comparison attempts to control for dispositional differences known to exist between women choosing midwifery and obstetrician care (Cordray & Davidson, 2004; van DerHulst et al., 2004; Howell-White, 1997; Hundley et al., 1997). The FBC-Midwife comparison also had differential risk profiles between groups. However, for this comparison, the FBC group was *less* likely to have demographic, behavioral, and medical risks than the Hospital group. Again, controlling for the different risk profiles, the FBC participants revealed better or equal outcomes relative to their hospital counterparts. Incidence of low birth weight, small for gestational age, preterm birth, fetal distress, and maternal complications are all significantly more likely to occur in the Midwife-attended Hospital group than the FBC group. Whereas, the incidences of low Apgar scores and abnormal infant conditions are both statistically equivalent between the two groups, only the odds of having a large weight for gestational age (LGA) infant were significantly greater in the FBC group than in the Hospital Group. for the FBC-Midwife analysis. Implications for having greater LGA infants are discussed below.

Although the design of midwife-attended hospital comparison groups helps to minimize the impact of selection bias, there always remains the question of whether these

outcomes would have been observed in a randomized study. Methodological researchers studying quasi-experimental design have developed tools that help observational studies come closer to estimating the effects of a randomized study (Rosenbaum, 2002). There is some concern that logistic regression alone cannot be completely trusted to control for all pre-existing group differences when there are many covariates and small proportions of outcomes (Cepeda et al., 2003; Rubin, 2001). Therefore, the analysis of FBC versus midwife-attended Hospital participants included a propensity score analysis to balance any remaining differences between the two groups.

The propensity score analysis for the FBC-Midwife comparison provided evidence that the pre-existing differences between groups were great enough to warrant concern for whether logistical regression models could adequately control for the full risk profile. The systematic check for balance between groups revealed that risks were not completely balanced between groups in the FBC-Midwife comparison. Therefore, an extended propensity score model, including interactions of risks, was created to balance the set of covariates between groups. After controlling for risks using the extended propensity score covariate, however, outcomes were not markedly different then when the logistic regression model was used to control for individual risk variables. The lack of change in odds ratio estimates provides statistical evidence that the more common use of logistic regression to control for risk profiles in this study only minimally biases the results.

Rubin (1997, 2001) postulates that if all covariates are balanced among groups using sub-classification on the propensity score, the technique mimics the role of a randomized design by eliminating 90% of bias based on any observed risk factor. By

evaluating the balance of risk profiles within quintiles in the FBC-Midwife comparison, the propensity score comes very close to balancing all risks and their interactions among groups. Therefore, according to propensity score theory, it is reasonable to believe that estimates for each outcome are approximating estimates from a randomized experiment.

Large for Gestational Age

Large weight for gestational age (LGA), was the only outcome that had a significantly greater likelihood of occurring in the FBC group than in the midwife-attended hospital group, even after controlling for the risk profiles (FBC: 0.127 v Hospital: 0.083; OR: 1.45, 95%1.26-1.67). Although there is a similar trend, the odds for LGA in the FBC-Low comparison are not statistically different between groups. The question that must be addressed is whether this outcome is an indicator of poor health and presents a problem for the safety of FBCs.

The definition of LGA is typically an infant that weighs more than 90 percent of all babies of the same gestational age or over 4,000 grams (8 pounds, 13 ounces). Researchers consider diabetes as the most common cause of LGA infants, but genetics and weight gain in pregnancy also contribute to large birth weight. LGA is a concern because if babies are large, women can have a difficult delivery possibly leading to birth injury or it may be a marker for undetected diabetes that leads to other problems (Scott, Gibbs, Karlan, & Haney, 2003). Both diabetes and a previous >4000-gram infant occur more often in the FBC group than in the Hospital group risk profiles. Therefore, it is not clear why LGA would remain a significant effect in the FBC-Midwife comparison after controlling for diabetes and previous pregnancies.

Only 29 LGA infants (0.093) in the FBC group were transferred to the hospital for delivery, indicating that there may not have been additional problems associated with large birth weight. The mean birth weights for LGA infants were found to be significantly different between the two groups (FBC: M = 4298.6 grams, SD: 253.86 v. Hospital: M = 4200.7 grams, SD: 232.64, t = -6.32, p <0.001). However, if there were negative problems associated with LGA, we would expect to see significant differences between groups on other complications within this sub-group. A supplementary assessment of only the LGA cases indicated that there was not a significant difference between FBC and Hospital groups on either incidence of maternal complications (FBC: 0.24 v. Hospital: 0.21; OR: 1.13, 95%CI: 0.83-1.53) or abnormal conditions of the infant (FBC: 0.087 v. Hospital: 0.069; OR: 1.28, 95%CI: 0.80-2.04). By these indicators, there was no evidence that the LGA cases in this sample posed a greater health threat to women using FBC services.

Comparing Outcomes Between FBC-Low and FBC-Midwife Analyses

The interpretation for all outcomes between the FBC-Low and FBC-Midwife analyses showed similar patterns of treatment effects. The purpose for using only midwife-attended births was to be able to closely match FBC care with a hospital comparison group on unmeasured characteristics. Based on this rationale, it was expected that the FBC-Midwife comparison would minimize selection bias attributed to these characteristics, and therefore provide a more precise estimate of treatment effects. Despite a different pattern of risks for the FBC group in the two analyses (i.e. FBC group in comparison to the Hospital group had more risks in FBC-Low and FBC had less risks

in the FBC-Midwife analyses), the majority of odds ratios in the FBC-Midwife still move closer to 1.0 (Table 16). This pattern of effects provided descriptive evidence that the FBC-Midwife groups were better matched on the set of risk factors that influence childbirth outcomes.

Table 16: Comparing Outcomes Between FBC-Low and FBC-Midwife Analyses

Outcomes	FBC-Low		FBC-Midwife	
	Adjusted Odds Ratio	95% Confidence Interval	Adjusted Odds Ratio	95% Confidence Interval
LBW (<2500)	0.39***	0.27-0.55	0.72*	0.54-0.98
Small for gestational age	0.70**	0.55-0.89	0.80*	0.66-0.96
Large for gestational age	1.21	0.99-1.47	1.47***	1.27-1.69
Preterm birth (<37)	0.36***	0.27-0.49	0.59***	0.46-0.76
Apgar score, <7 at 1	0.85	0.67-1.09	1.07	0.89-1.30
Apgar score, <7 at 5	0.87	0.42-1.80	1.13	0.65-1.96
Fetal distress	0.24***	0.18-0.33	0.31***	0.24-0.40
Incidence of abnormal conditions of the infant	0.44***	0.35-0.54	0.99	0.82-1.20
Incidence of maternal complications during labor	0.38***	0.33-0.44	0.84**	0.75-0.95

* p<0.05; ** p<0.01; *** p<0.001

The pattern of outcome across both analyses provided evidence that FBCs do not have more adverse outcomes than Hospitals and can therefore be considered a safe service for childbirth. In fact, the outcomes were very similar to previous studies that compared FBC with Hospital care using different methodological designs. The Jackson et al. (2003) study prospectively observed differences between low-income women using FBC and Hospital services. Similarly to the FBC-Low analysis, the hospital comparison group was sampled based on eligibility criteria regarding level of risks. However, the Jackson et al (2003) study was able to measure risk more precisely with the use of birth

center eligibility criteria and the use of retrospective chart reviews. They were also able to use a full intent-to-treat analysis plan that captured all transfer cases. Jackson et al. (2003) found no difference between the groups on major maternal or neonatal complications including LBW, small for gestational age, large for gestational age, preterm birth, and low Apgar scores. Even with the more precise risk assessment than what birth certificate data assessment allowed, however, the results from this study supported the findings from the FBC-Low comparison.

Use of midwife-attended hospital births as the comparison group in the FBC-Midwife analysis was a direct extension of the Kellogg Birth Center Evaluation project (Cordray & Davidson, 2004). Their evaluation design used a much smaller sample of FBC births and midwife-attended hospital controls. They reported a pattern of outcomes similar to the findings in the current study. Over four years (1999-2002), the evaluation found that the majority of outcomes were statistically equivalent between the three Tennessee birth centers and their matched group of midwife-attended hospital births. Only one outcome, an aggregated labor and delivery complication variable, was significantly more likely to occur in the Hospital group than in two of the three FBC groups. The FBCs in the Kellogg Evaluation were more heterogeneous with respect to their organizational stability, which may explain the differences in outcomes. The evaluation did not measure large for gestational age (LGA); therefore, it is unknown whether the outcome stayed consistent with findings from the FBC-Midwife analysis. The expanded number of years covered in the current study (1990-2002) shows consistent findings in FBC outcomes as compared with the results from the evaluation project that covered a fewer number of years (1999-2002). The consistent outcomes

across twelve years seems to provide more evidence that the low rates of poor outcomes are dependable.

The use of propensity analysis to control for risk profiles in the FBC-Midwife analysis theoretically mimicked the randomization process by balancing observed risk factors across both groups. Since the use of propensity scores cannot balance the unobserved differences between groups, by limiting the propensity analysis to only Midwife-attended Hospital groups, at least a portion of the unobserved differences related to choice of care are controlled. Although the outcome estimates do not change significantly between statistically controlling for the propensity score versus controlling for the individual risks, it is important to note that the findings for this study are also consistent with a randomized control trial conducted in Stockholm, Sweden, which assessed in-hospital birth centers. Waldenstrom et al. (1997) compared an in-hospital birth center with standard maternity care. The same midwives cared for women in the in-hospital birth center with physician care only if medical transfer occurred, whereas a team of midwives and physicians cared for women in the standard maternity care. Using an intent-to-treat analysis, Waldenstrom et al (1997) found no statistical differences in perinatal mortality, major maternal morbidity, or neonatal morbidity. However, the authors were inconclusive about whether the in-hospital birth centers were as safe as the standard center because of a small sample size (birth center, n = 928 v. standard, n = 932) that did not have enough statistical power to identify small differences.

The combination of using Midwife-attended Hospital groups for controlling for dispositional characteristics, with the use of the propensity score analysis to make up for limitations seen in regression models, produced estimates of treatment effects that have

minimal bias. The goal of this study was to evaluate the safety of FBCs as compared to hospital births. By employing methodological procedures that addressed limitations in previous research, the results from both FBC-Low and FBC-Midwife analyses strengthened the evidence supporting the safety of FBCs in general. Nevertheless, there are potential limitations to the design of this study that need to be addressed.

Potential Limitations

Two potential limitations to interpreting safety of FBCs in this study need further discussion. The first is the inability to capture outcomes from all women who were transferred from the FBC to the hospital during labor. The second is the inability to include all important biopsychosocial risks for poor pregnancy outcomes in the complete risk profiles.

Potential Limitations Due to Underestimating Rates of Transfer to the Hospital

The most critical time point for care at FBCs is during labor and delivery. How care is handled during unexpected problems is essential to whether FBC service can be considered safe for childbirth. In order to address this question as part of the evaluation on safety, women who were transferred during labor must be included in the treatment group. Otherwise, only the healthiest of an already low risk group would be used to evaluate the FBC service. As prescribed by the intent-to-treat model of analysis, the FBC sample included women who delivered at the hospital after being transferred. Although the birth certificate has a variable that indicated whether a woman had been transferred from another institution, birth certificate records are notorious for underestimating the

true number of cases. The extent to which the birth certificate data accurately reports the intrapartum transfer rates, however, can be estimated with what is known from the results of the Kellogg Birth Center Evaluation (Cordray & Davidson, 2004).

Comparing the rate of transfers in the current data set with the rate of transfers reported in the Kellogg Birth Center Evaluation revealed that the underreporting of transfers on the birth certificate was not a large problem. The evaluation collected medical chart data, between July 2001 and October 2003, at the same urban-based, birth center included as one of the centers in the current study. According to the medical chart review at this FBC, 44 out of 208 (0.212) women transferred for medical or personal reasons during the intrapartum stage of labor. For approximately the same time period (fiscal years 2001 and 2002), the birth certificate record indicated that there were 31 out of 210 (0.148) women that were transferred from the FBC and delivered in the hospital. It can be estimated, therefore, that the birth certificate may underestimate the true intrapartum transfer rate by a proportion of approximately 0.064. This translates into approximately 14 additional women for the urban-based birth center. The proportional rate of intrapartum transfer is also consistent with both the Jackson et al. (2003) study (0.185) and the Waldenstrom et al. (1997) study (0.19).

It can be assumed that any woman who is transferred from the FBC during labor has medical risk for poor outcomes; therefore, inclusion of these cases in the overall comparison of the FBC and Hospital group actually weights the FBC group towards having poorer outcomes. Unfortunately, this data set does not allow for an analysis that compares transfers from the FBC group to those in the midwife-attended Hospital group with similar problems. Women in the hospital group with similar problems would

typically remain at the same hospital location, but could have been transferred internally to physician care. Despite the inclusion of women who were transferred, the FBC group still had equivalent or better outcomes as the Hospital group. And although this dataset does not include all of the transfer cases in the FBC group, the underestimate likely would not change interpretation of the treatment effect.

Potential Limitations Due to Unmeasured Risk Factors

There is always a possibility that additional risk factors not measured in this study would impact the treatment effect between FBC and Hospital services. Despite matching for both midwifery care and statistically adjusting for a substantial number of risks, there are important risk factors known to predict poor pregnancy outcomes that are left out of the model. Research has show that nutritional status, rates of infections, psychological factors, and overall health problems related to factors of poverty account for some of the variance in poor pregnancy outcomes (Dunkel-Schetter, Gurung, Lobel & Wadhwa, 2001; Taylor, Repetti, & Seeman, 1997; Zimmer-Gembeck & Helfand, 1996; Paneth, 1995; Kramer, 1987). However, it is important to note that the extent to which these factors impact outcomes is also dependent on screens during prenatal care. In the FBC-Midwife comparison, the expectation is that philosophies of care between the two groups are comparable. If these risks were measured, we would expect that the midwives practicing at the hospital would handle them in the same way as the midwives practicing at the FBC. Furthermore, even if these could be measured, it is very unlikely that further adjustment for these factors would make the FBC group look less safe than the Hospitals. At best, inclusion of these additional factors would most likely continue to close the gap between

the two services. The safety of FBCs, being the primary research question, would not change.

Concluding Statements

This is the first large assessment of FBC safety in Tennessee that is an extension to the Kellogg Birth Center Evaluation Project (Cordray and Davidson, 2004). The overall results of this study provide strong evidence that freestanding birth centers are doing their job well. Those of us persuaded by empirical evidence should find comfort in the repeated findings that FBCs do not have higher rates of poor childbirth outcomes than Hospital services. The low rates of poor outcomes indicate that the providers practicing at FBCs effectively manage the risk status of their clients and provide a healthy service for women and infants during pregnancy and childbirth.

Critics of previous FBC research have suggested that self-selection bias makes interpretation of outcomes difficult. Close assessment of risk profiles between groups in this study, confirm that their concerns are well founded. FBC and Hospital groups in both analyses had statistically significant differences on pre-existing risks. After paying careful attention to balancing pre-existing differences in risk profiles, the interpretation of FBC care remains the same. FBCs have low rates of poor outcomes and FBCs are just as safe as Hospital services.

The question of whether FBCs are safe has impeded insurance negotiations, organizational support, and professional practice rights associated with local maternal health systems. Major obstetrical texts (Cunningham et al., 1997) and professional organizations repeatedly state that the only safe place for childbirth is in the hospital.

This statement is certainly true for women at risk for medical problems; however, findings from this study provide empirical support that women who meet FBC eligibility criteria and remain at low risk throughout their pregnancy also have the option of using a FBC service.

The empirical basis for such statements against the use of FBCs likely develops from the knowledge that birth is uncertain. The Institute of Medicine estimates that there are 20% false positive assignments to initial low risk status and 14% false positive assignments to the high-risk status for childbirth (IOM, 1982). Moreover, given the transfer rates at FBCs in the current study and in previous studies, this concern is reasonable. However, to use the argument of uncertainty as the ultimate reason why FBCs can never be a safe option is unfounded. Evaluating risk for all pregnancies is an ongoing effort. Professionals learn to identify these risks and professionals working at accredited FBCs are trained to identify clinical indicators when transfer to the hospital is needed. If there were no cases transferred from the FBC to the Hospital, then FBC providers would not be following through with an ongoing risk assessment. The inclusion of these cases in the overall assessment of safety provides more evidence that the FBCs are providing overall safe care for their clients.

The acknowledgement of FBCs as a safe option in the larger maternal health system depends on the question of what constitutes good evidence. A randomized control study would obviously produce an unbiased estimate of the treatment effects between FBC and Hospital care. However, ethical considerations between what researchers need and the needs of individuals limit this possibility. At this time, observational studies are the most effective way of studying the effect of FBC service.

Researchers focusing specifically on improving causal modeling in observational studies have developed tools to help make evidence from observational studies stronger (Rosenbaum, 2002). In this study, the research design attempts to couple some of these methods with appropriate selection of comparison groups to come as close as possible to generating the type of evidence a randomized design would produce. Although there always remains room for improvement, the results from this study are consistent with research comparing the outcomes of FBC and Hospital services for over two decades.

This study expands on previous research assessing the safety of freestanding birth centers by utilizing strategies to minimize the effect of bias often found in non-randomized research designs. Results from this study contribute to the overall understanding of the efficacy of FBCs, as well as methods for minimizing bias in observational studies in this area. With pending crises in the maternal health care system, largely over access to care, credible options are important for public health planning. To be a viable service in the maternal health system, safety is paramount. The results from this study empirically support the use of freestanding birth centers as a safe option for childbirth.

APPENDIX A

VARIABLE DESCRIPTIONS

Table A-1: Descriptions of biological risks

Biological Risks	Descriptions of variables (State of Tennessee, 2000*)
Risks predating pregnancy	
Parity	Total number of pregnancies
Previous infant 4000+ grams	The birth weight of previous live-born child was over 4,000 grams.
Prior history of small-for-gestational age	Previous birth of an infant prior to term usually considered earlier than 37 completed weeks of gestation. Fetal deaths not included.
Prior history of preterm birth	Previous birth of an infant weighing less than the tenth percentile for gestational age using a standard weight for age chart. No fetal deaths.
Number of other terminations	Includes spontaneous miscarriages, spontaneous abortions, induced terminations, or pregnancies that resulted in fetal death (stillborn) at anytime after conception.
Anemia	Hematocrit of less than 30% or Hemoglobin less than 10.0 g/dl during pregnancy. A symptom of some underlying disease, such as iron deficiency, chronic blood loss, sickle cell anemia.
Cardiac disease	Mother has diagnosis of a disease of the heart, such as rheumatic heart disease, congenital heart disease, cyanotic heart disease, coronary thrombosis, bacterial endocarditis, cardiomyopathy, mitral valve proplapse, cardiovascular complications from Marfan syndrome, coarctation of the aorta, or kyphoscoliotic heart disease during this pregnancy
Acute or chronic lung disease	Mother has diagnosis of a disease of the lungs during this pregnancy. Acute is a short and sharp course of lung disease like pneumonia or acute bronchitis. Chronic is of long duration, denoting a disease of slow progress and long continuance, like tuberculosis, cystic fibrosis, chronic bronchitis, chronic obstructive bronchitis, pulmonary edema, chronic obstructive emphysema, persistent asthma, or chronic asthmatic bronchitis.
Diabetes	Mother has diagnosis of type 1, juvenile onset diabetes, type 2, adult onset diabetes, or gestational diabetes mellitus during this pregnancy.
Active genital herpes	Infection of the skin of the genital area by herpes simplex virus.
Hemoglobinopathy	A hematologic disorder caused by alteration in the genetically determined molecular structure of hemoglobin, which results in a characteristic complex of clinical and laboratory abnormalities.
Hypertension, chronic	Blood pressure persistently greater than 140/90, diagnosed prior to the onset of the pregnancy or before the 20th week of gestation.
Renal disease	Mother has diagnosis of a kidney disease, such as, acute or chronic pyelonephritis, glomerulonephritis, nephrosis, acute tubular necrosis, renal cortical necrosis, obstructive renal failure, diabetic nephropathy or polycysitic kidney disease during this pregnancy.
Other	Other medical risk factors experience by the mother that may cause or contribute to complications of this pregnancy. Examples are AIDS, preeclampsia, rubella, syphilis, gonorrhea, early onset of delivery and mental disorder.

Table A-1 continued

Biological risks	**Descriptions of variables (State of Tennessee, 2000*)**
Risks in current pregnancy	
Hydramnios or polyhydramnios	Excessive volume of amniotic fluid, somewhat arbitrarily defined as greater than 2,000 ml. Hydramnios sufficient to cause clinical symptoms (usually > 3,000 ml) occurs in about 1 in 1,000 pregnancies excluding multifetal pregnancies.
Oligohydramnios	Volume of amniotic fluid falls or is far below normal, sometimes with a few ml of viscid fluid. Cause is not understood.
Hypertension, pregnancy-induced	An increase in blood pressure of at least 30 mm Hg systolic or 15 mm Hg diastolic on two measurements taken 6 hours apart after the 20th week of gestation.
Eclampsia	The occurrence of convulsions and/or coma unrelated to other cerebral conditions in women with signs and symptoms of preeclampsia. Occurs in neglected or, less often, fulminate cases of pregnancy-induced hypertension. Seizures are of grand mal type and may first appear before labor, during labor, or up to 48 hours postpartum.
Incompetent cervix	Characterized by painless dilation of the cervix in the second trimester or early in the third trimester of pregnancy with prolapse of membranes through the cervix and ballooning of the membranes into the vagina, followed by rupture of the membranes and subsequent expulsion of the fetus.
Rh sensitization	The process or state of becoming sensitized to the Rh factors (i.e., Rh antigen(s), especially D antigen) as when an Rh-negative woman is pregnant with an Rh-positive fetus. Unless mother was previous sensitized by transfusion, a first pregnancy is rarely affected. Preventative measures such as the use of Rhogam, are not included.
Uterine bleeding	Any clinically significant bleeding during the pregnancy taking into consideration of the gestational age of the patient. Any 2nd or 3rd trimester bleeding of the uterus prior to the onset of labor.

* Descriptions taken from the *Handbook on birth registries, fetal deaths, and induced terminations of pregnancy* (September, 2001). State of Tennessee: Department of Health and Health Information, Office of Vital Records.

Table A-2: Proportions of risks in FBC group and two comparison groups for aggregated medical risk variable proportions

De-aggregated medical risks	FBC group N= 2463 #(proportion)	Low risk Hospital group N = 2500 #(proportion)	Midwife Hospital group N = 11653 #(proportion)
Anemia	85 (0.0346)	0	848 (0.0729)
Cardiac disease	11 (0.0045)	0	70 (0.0060)
Acute or chronic lung disease	15 (0.0061)	0	135 (0.0116)
Diabetes	52 (0.0212)	0	197 (0.0169)
Active genital herpes	3 (0.0012)	1005 (0.0067)	42 (0.0036)
Hydraminos	4 (0.0016)	551 (0.0037)	31 (0.0027)
Oligohydraminos	15 (0.0061)	0	233 (0.020)
Hemoglobinopathy	1 (0.0004)	31 (0.0002)	0
Chronic hypertension	3 (0.0012)	0	70 (0.0060)
Pregnancy induced hypertension	33 (0.0134)	0	433 (0.0372)
Eclampsia	0	0	9 (0.0008)
Incompetent cervix	3 (0.0012)	0	9 (0.0008)
Renal disease	0	0	32 (0.0027)
Rh sensitization	5 (0.0020)	393 (0.0026)	33 (0.0028)
Uterine bleeding	7 (0.0028)	0	52 (0.045)

Table A-3: Descriptions of outcomes

Infant Outcomes	**Descriptions of variables (State of Tennessee, 2000*)**
Low birth weight	<2500 grams
Large for gestational age	Infants with birth weight in the 90th percentile for gestational age following the United States national reference for fetal growth (Alexander, Himes, Kaufman, Mor, & Kogan (1996)
Small for gestational age	Infants with birth weight below the 10th percentile following the United States national reference for fetal growth (Alexander et al.(1996)
Preterm birth	< 37 weeks gestation
Apgar score (< 7 at 1 minute; < 7 at 5 minutes)	A scoring system used to evaluate newborns at one minute and five minutes after delivery. The total score is achieved by assessing five signs: heart rate, respiratory effort, muscle tone, reflex irritability, and color. Each of the signs is assigned a score of 0, 1, or 2. The highest possible score is 10. Less than 7 is considered potentially a risk marker.
Fetal distress	Signs indicating fetal hypoxia (deficiency in amount of oxygen research fetal tissues)
Incidence of abnormal conditions	This aggregated variable includes the following conditions: Anemia; Birth injury; Fetal alcohol syndrome; Fetal drug syndrome; Hyaline membrane disease/RDS; Maconuim aspiration syndrome; Assisted ventilation < 30 minutes; Assisted ventilation > 30 minutes; Seizures, and Other.
	Descriptions of individual abnormal conditions
Anemia	Hemoglobin level of less than 13.0 g/dL or a hematocrit of less than 39%
Birth injury	Impairment of the infant's body function or structure due to adverse influence that occurred at birth
Fetal alcohol syndrome	A syndrome of altered prenatal growth and development occurring in infants born to women who consumed excessive amounts of alcohol during pregnancy
Hyaline membrane disease/RDS	A disorder primarily of prematurity, manifested clinically by respiratory distress and pathologically by pulmonary hyaline membranes and incomplete expansion of the lungs at birth.
Meconium aspiration syndrome	Aspiration of meconium by the fetus or newborn affecting the lower respiratory system.
Assisted ventilation (< 30 min; >30 min)	A mechanical method of assisting respiration for newborns with respiratory failure
Seizures	A seizure of any etiology
Maternal Outcomes	
Number of maternal complications during labor	The aggregated variable includes the following complications: Febril (>100 degrees F); Maconium, moderate/heavy; Premature rupture of membranes (>12 hours and < 24 hours); Prolonged rupture of membranes (>24 hours); Abruptio placenta; Placenta previa; Other excessive bleeding; Seizures during labor; Precipitous labor (< 3 hours); Prolonged labor (> 20 hours); Dysfunctional labor; Breech; Malpresentation; Cephalepelvic disproportion; Cord prolapse; Anesthetic complications; Other
	Descriptions of maternal complications
Febril (>100 degrees F)	Fever occurring during labor and/or delivery
Meconium, moderate/heavy	Meconium consists of undigested debris from swallowed amniotic fluid, various products of secretion, excretion, and shedding by the gastrointestinal tract; moderate to heavy amounts of meconium in the

	amniotic fluid noted during labor and /or delivery.
Premature or prolonged rupture of membranes	Rupture of the membranes at any time during pregnancy and more than 12 hours before the onset of labor;
Precipitous or prolonged labor	Extremely rapid labor (lasting less than 3 hours) or abnormally slow progress of labor (lasting more than 20 hours)
Abruptio placenta	Premature separation of a normally implanted placenta from the uterus
Placenta previa	Implantation of the placenta over or near the internal opening of the cervix
Other excessive bleeding	The loss of a significant amount of blood from conditions other than abruption placenta or placenta previa
Seizures during labor	Maternal seizures occurring during labor from any cause
Dysfunctional labor	Failure to progress in a normal pattern of labor
Cephalopelvic disproportion	The relationship of the size, presentation, and position of the fetal head to the maternal pelvis that prevents dilation of the cervix and /or descent of the fetal head.
Cord prolapse	Premature expulsion of the umbilical cord in labor before the fetus is delivered
Anesthetic complications	Any complication during labor and /or delivery brought on by an anesthetic agent or agents

* Descriptions taken from the *Handbook on birth registries, fetal deaths, and induced terminations of pregnancy* (September, 2001). State of Tennessee: Department of Health and Health Information, Office of Vital Records.

APPENDIX B

PROPENSITY SCORE ANALYSES

Table B-1: Residual diagnostics for the first propensity score model (including only original 16 risk factors) between Hospital and FBC group to check overall balance on each risk variable.

Risk	N	Mean	Std Dev	Variance	Ratio
<19 years old	11475	3.507E-06	0.319	0.102	1.225
	2366	-4.946E-05	0.289	0.083	
>35 years old	11475	1.795E-04	0.188	0.035	0.705
	2366	-8.709E-04	0.223	0.050	
<12 years education	11475	-2.896E-05	0.430	0.185	1.130
	2366	1.379E-04	0.405	0.164	
African American risk marker	11475	1.606E-05	0.308	0.095	1.291
	2366	-7.788E-05	0.271	0.074	
Unmarried	11475	-3.747E-05	0.444	0.197	1.062
	2366	1.805E-04	0.431	0.186	
>20 weeks (or 5 months) gestation at first prenatal care visit	11475	6.728E-06	0.247	0.061	1.300
	2366	-6.034E-05	0.216	0.047	
Cigarette smoking	11475	3.174E-05	0.395	0.156	0.960
	2366	-2.179E-04	0.403	0.162	
Drinking alcohol	11475	3.515E-07	0.084	0.007	1.191
	2366	-2.166E-06	0.077	0.006	
Use of drugs	11475	2.910E-05	0.101	0.010	0.786
	2366	-1.412E-04	0.114	0.013	
First pregnancy - parity risk	11475	-5.651E-05	0.466	0.218	1.139
	2366	2.725E-04	0.437	0.191	
Prev. infant 4000+ grams	11475	2.131E-04	0.101	0.010	0.509
	2366	-1.034E-03	0.142	0.020	

Prev. small for gestation	11475	4.935E-05	0.058	0.003	0.631
	2366	-2.414E-04	0.073	0.005	
Prev. preterm	11475	1.002E-04	0.114	0.013	0.730
	2366	-4.999E-04	0.133	0.018	
Greater than 3 previous terminations - (miscarriage, abortion, still birth)	11475	7.825E-05	0.067	0.004	0.566
	2366	-3.795E-04	0.089	0.008	
Medical risk present	11475	-1.798E-04	0.281	0.079	1.671
	2366	8.661E-04	0.218	0.047	
Other	11475	-1.238E-04	0.205	0.042	2.130
	2366	6.004E-04	0.141	0.020	

Table B-2: Variables used to calculate the expanded propensity score model (including interactions)

Variables used in the Propensity Model	Description
agrisk_y	<19 years old
agrisk_o	>35 years old
edurisk	<12 years education
racerisk	Race (African American)
nmarrrisk	Unmarried
pncare_risk	> 20 weeks (5 months) gestation at first prenatal care visit
cigrisk	Cigarette smoking
alcrisk	Drinking alcohol
drugrisk	Use of drugs
parrisk2	First pregnancy
MR13_3	Previous infant >4000 grams
MR15_3	Previous small for gestation
MR14_3	Previous preterm
prterminate3	>2 previous terminations
summed2	Any incidence of medical risk
MR19	Other category for medical risk
racecig	racerisk*cigrisk
educig	edurisk*cigrisk
cigmar	cigrisk*nmarrrisk
cigpn	cigrisk*pncare_risk
Alcold	agrisk_o*alcrisk
alcrace	racerisk*alcrisk
alcmar	alcrisk*nmarrrisk
alcpn	alcrisk*pncare_risk
alccig	alcrisk*cigrisk
drugyoung	drugrisk*agrisk_y
drugedu	drugrisk*edurisk
drugrace	drugrisk*racerisk
drugmar	drugrisk*nmarrrisk
drugpn	drugrisk*pncare_risk
drugcig	drugrisk*cigrisk
drugalc	drugrisk*alcrisk
youngparr	agrisk_y*parrisk2
oldparr	agrisk_o*parrisk2
eduparr	edurisk*parrisk2
raceparr	racerisk*parrisk2
marparr	nmarrrisk*parrisk2
cigparr	cigrisk*parrisk2
drugparr	drugrisk*parrisk2
racemr19	racerisk*mr19
marrmr19	nmarrrisk*mr19
pnmr19	pncare_risk*mr19
parmr19	parrisk2*mr19
drugmr19	drugrisk*mr19
youngsummed2	agrisk_y*summed2
edusummed2	edurisk*summed2
racesummed2	racerisk*summed2
marsummed2	nmarrrisk*summed2

cigsummed2	cigrisk*summed2
alcsummed2	alcrisk*summed2
drugsummed2	drugrisk*summed2
parsummed2	parrisk2*summed2
mr19summed2	mr19*summed2
youngedu	agrisk_y*edurisk
oldedu	agrisk_o*edurisk
youngrace	agrisk_y*racerisk
oldrace	agrisk_o*racerisk
raceedu	racerisk*edurisk
youngmarr	agrisk_y*nmarrrisk
oldmarr	agrisk_o*nmarrrisk
edumarr	edurisk*nmarrrisk
racemarr	racerisk*nmarrrisk
youngpn	agrisk_y*pncare_risk
oldpn	agrisk_o*pncare_risk
edupn	edurisk*pncare_risk
racepn	racerisk*pncare_risk
marpn	pncare_risk*nmarrrisk
youngcig	agrisk_y*cigrisk
oldcig	agrisk_o*cigrisk
Interactions not significantly different between groups, and therefore not included in the propensity score	
Alcohol*<19 years old Alcohol*< 12 years education Use of drugs*>35 years old First pregnancy*late prenatal care First pregnancy*alcohol Summed2*>35 years old Summed2*late prenatal care MR19(medical risk other)*<19 years old MR19(medical risk other)*>35 years old MR19(medical risk other)*<12 years education MR19(medical risk other)*cigarette smoking MR19(medical risk other)*alcohol	

Table B-3: Ratios of residual variance between groups for extended propensity model

Risks		N	Mean	Std Dev	Variance	Ratio
<19 years old	Hosp	11475	0.000019	0.318634	0.101528	1.201706
	FBC	2366	-0.000106	0.290665	0.084486	
>35 years old	Hosp	11475	0.000317	0.210517	0.044317	0.613284
	FBC	2366	-0.001538	0.268816	0.072262	
<12 years education	Hosp	11475	-0.000008	0.431282	0.186004	1.061837
	FBC	2366	0.000037	0.418535	0.175172	
African American risk marker	Hosp	11475	0.000027	0.306774	0.094110	1.199560
	FBC	2366	-0.000160	0.280096	0.078454	
Unmarried	Hosp	11475	-0.000019	0.446899	0.199718	1.029507
	FBC	2366	0.000089	0.440448	0.193994	
>20 weeks (or 5 months) gestation at first prenatal care visit	Hosp	11475	0.000012	0.244706	0.059881	1.231671
	FBC	2366	-0.000067	0.220495	0.048618	
Cigarette smoking	Hosp	11475	0.000036	0.395535	0.156448	0.960564
	FBC	2366	-0.000200	0.403573	0.162871	
Drinking alcohol	Hosp	11475	0.000000	0.083599	0.006989	1.184237
	FBC	2366	-0.000002	0.076822	0.005902	
Use of drugs	Hosp	11475	0.000017	0.101050	0.010211	0.810815
	FBC	2366	-0.000083	0.112221	0.012594	
First pregnancy - parity risk	Hosp	11475	-0.000030	0.471615	0.222421	1.089301
	FBC	2366	0.000145	0.451870	0.204187	
Prev infant 4000+ grams	Hosp	11475	0.000179	0.118043	0.013934	0.465861
	FBC	2366	-0.000867	0.172947	0.029911	
Prev small for gestation	Hosp	11475	0.000018	0.058186	0.003386	0.631172
	FBC	2366	-0.000089	0.073239	0.005364	
Prev preterm	Hosp	11475	0.000057	0.114361	0.013078	0.657482
	FBC	2366	-0.000280	0.141037	0.019892	
Greater than 3 previous terminations - (miscarriage, abortion, still birth)	Hosp	11475	0.000030	0.067577	0.004567	0.543066
	FBC	2366	-0.000149	0.091701	0.008409	
Summed2	Hosp	11475	0.000187	0.318469	0.101423	1.594733
	FBC	2366	-0.000907	0.252188	0.063599	
Other risk	Hosp	11475	0.000177	0.259227	0.067199	2.062653
	FBC	2366	-0.000859	0.180496	0.032579	
youngedu = agrisk_y*edurisk;	Hosp	11475	0.000024	0.279527	0.078136	1.268058
	FBC	2366	-0.000148	0.248230	0.061618	
oldedu = agrisk_o*edurisk;	Hosp	11475	-0.000001	0.063860	0.004078	1.074953
	FBC	2366	0.000001	0.061593	0.003794	
youngrace = agrisk_y*racerisk;	Hosp	11475	0.000044	0.145410	0.021144	2.185452

	FBC	2366	-0.000215	0.098361	0.009675	
oldrace = agrisk_o*racerisk;	Hosp	11475	0.000006	0.057418	0.003297	0.782900
	FBC	2366	-0.000029	0.064893	0.004211	
raceedu = racerisk*edurisk;	Hosp	11475	0.000010	0.205221	0.042116	1.205482
	FBC	2366	-0.000055	0.186914	0.034937	
youngmarr = agrisk_y*nmarrrisk;	Hosp	11475	0.000089	0.267643	0.071633	1.662355
	FBC	2366	-0.000439	0.207584	0.043091	
oldmarr = agrisk_o*nmarrrisk;	Hosp	11475	0.000036	0.071296	0.005083	0.558466
	FBC	2366	-0.000180	0.095404	0.009102	
edumarr = edurisk*nmarrrisk;	Hosp	11475	0.000096	0.336050	0.112929	1.370328
	FBC	2366	-0.000468	0.287072	0.082410	
racemarr = racerisk*nmarrrisk;	Hosp	11475	0.000113	0.255817	0.065442	1.785196
	FBC	2366	-0.000547	0.191464	0.036658	
youngpn = agrisk_y*pncare_risk;	Hosp	11475	0.000010	0.115050	0.013237	1.514433
	FBC	2366	-0.000049	0.093489	0.008740	
oldpn = agrisk_o*pncare_risk;	Hosp	11475	-0.000001	0.023592	0.000557	n/a
	FBC	2366	0.000000	0.000000	0.000000	
edupn = edurisk*pncare_risk;	Hosp	11475	-0.000002	0.168898	0.028526	1.093953
	FBC	2366	0.000008	0.161482	0.026077	
racepn = racerisk*pncare_risk;	Hosp	11475	0.000024	0.109992	0.012098	0.843675
	FBC	2366	-0.000118	0.119749	0.014340	
marpn = pncare_risk*nmarrrisk;	Hosp	11475	0.000033	0.176160	0.031032	1.634014
	FBC	2366	-0.000163	0.137809	0.018991	
youngcig = agrisk_y*cigrisk;	Hosp	11475	-0.000002	0.167312	0.027993	1.046058
	FBC	2366	0.000009	0.163587	0.026761	
oldcig = agrisk_o*cigrisk;	Hosp	11475	0.000046	0.076454	0.005845	0.519004
	FBC	2366	-0.000221	0.106125	0.011262	
racecig = racerisk*cigrisk;	Hosp	11475	0.000014	0.109633	0.012019	1.709423
	FBC	2366	-0.000068	0.083853	0.007031	
educig =edurisk*cigrisk;	Hosp	11475	-0.000005	0.299740	0.089844	1.107043
	FBC	2366	0.000025	0.284881	0.081157	
cigmar =cigrisk*nmarrrisk;	Hosp	11475	-0.000008	0.290582	0.084438	1.045083
	FBC	2366	0.000030	0.284245	0.080795	
cigpn =cigrisk*pncare_risk;	Hosp	11475	0.000021	0.140852	0.019839	1.672153
	FBC	2366	-0.000104	0.108924	0.011864	

alcold = agrisk_o*alcrisk;	Hosp	11475	0.000000	0.030947	0.000958	1.133830
	FBC	2366	0.000000	0.029063	0.000845	
alcrace = racerisk*alcrisk;	Hosp	11475	0.000005	0.037989	0.001443	3.421601
	FBC	2366	-0.000022	0.020537	0.000422	
alcmar = alcrisk*nmarrrisk;	Hosp	11475	0.000000	0.065162	0.004246	1.116972
	FBC	2366	0.000001	0.061656	0.003801	
alcpn = alcrisk*pncare_risk;	Hosp	11475	-0.000001	0.024870	0.000619	n/a
	FBC	2366	0.000000	0.000000	0.000000	
alccig = alcrisk*cigrisk;	Hosp	11475	0.000003	0.065674	0.004313	1.461174
	FBC	2366	-0.000013	0.054331	0.002952	
drugyoung = drugrisk*agrisk_y;	Hosp	11475	0.000000	0.046627	0.002174	1.030594
	FBC	2366	0.000001	0.045930	0.002110	
drugedu = drugrisk*edurisk;	Hosp	11475	0.000007	0.072910	0.005316	1.830623
	FBC	2366	-0.000034	0.053888	0.002904	
drugrace = drugrisk*racerisk;	Hosp	11475	0.000002	0.046405	0.002153	1.702750
	FBC	2366	-0.000012	0.035562	0.001265	
drugmar = drugrisk*nmarrrisk;	Hosp	11475	0.000005	0.088181	0.007776	0.888512
	FBC	2366	-0.000022	0.093550	0.008752	
drugpn = drugrisk*pncare_risk;	Hosp	11475	0.000000	0.034912	0.001219	0.963106
	FBC	2366	-0.000001	0.035574	0.001266	
drugcig = drugrisk*cigrisk;	Hosp	11475	-0.000001	0.085217	0.007262	1.079726
	FBC	2366	0.000002	0.082011	0.006726	
drugalc = drugrisk*alcrisk;	Hosp	11475	0.000000	0.044704	0.001998	1.182173
	FBC	2366	0.000000	0.041116	0.001691	
youngparr = agrisk_y*parrisk2;	Hosp	11475	0.000002	0.292111	0.085329	1.156666
	FBC	2366	-0.000011	0.271609	0.073771	
oldparr = agrisk_o*parrisk2;	Hosp	11475	0.000030	0.070930	0.005031	0.562158
	FBC	2366	-0.000145	0.094602	0.008950	
eduparr = edurisk*parrisk2;	Hosp	11475	-0.000003	0.308439	0.095134	1.118607
	FBC	2366	0.000013	0.291629	0.085047	
raceparr = racerisk*parrisk2;	Hosp	11475	-0.000003	0.189560	0.035933	1.081982
	FBC	2366	0.000012	0.182237	0.033210	
marparr = nmarrrisk*parrisk2;	Hosp	11475	0.000025	0.352533	0.124280	1.248637
	FBC	2366	-0.000138	0.315487	0.099532	
cigparr = cigrisk*parrisk2;	Hosp	11475	-0.000007	0.240221	0.057706	1.063075

	FBC	2366	0.000018	0.232985	0.054282	
drugparr = drugrisk*parrisk2;	Hosp	11475	0.000030	0.070118	0.004916	0.567962
	FBC	2366	-0.000151	0.093040	0.008656	
racemr19 = racerisk*mr19;	Hosp	11475	0.000065	0.142397	0.020277	2.665735
	FBC	2366	-0.000316	0.087215	0.007607	
marrmr19 = nmarrrisk*mr19;	Hosp	11475	0.000095	0.179429	0.032195	2.497934
	FBC	2366	-0.000467	0.113528	0.012889	
pnmr19 = pncare_risk*mr19;	Hosp	11475	0.000009	0.068417	0.004681	2.233058
	FBC	2366	-0.000045	0.045784	0.002096	
parmr19 = parrisk2*mr19;	Hosp	11475	0.000033	0.181690	0.033011	1.658880
	FBC	2366	-0.000161	0.141066	0.019900	
drugmr19 = drugrisk*mr19;	Hosp	11475	0.000000	0.043734	0.001913	1.131656
	FBC	2366	0.000000	0.041111	0.001690	
youngsummed2 = agrisk_y*summed2;	Hosp	11475	0.000074	0.136818	0.018719	3.711507
	FBC	2366	-0.000358	0.071018	0.005044	
edusummed2 = edurisk*summed2;	Hosp	11475	0.000145	0.181395	0.032904	3.241134
	FBC	2366	-0.000703	0.100757	0.010152	
racesummed2 = racerisk*summed2;	Hosp	11475	0.000036	0.146828	0.021558	1.922324
	FBC	2366	-0.000173	0.105900	0.011215	
marsummed2 = nmarrrisk*summed2;	Hosp	11475	0.000126	0.200075	0.040030	2.428170
	FBC	2366	-0.000625	0.128396	0.016486	
cigsummed2 = cigrisk*summed2;	Hosp	11475	0.000072	0.174385	0.030410	2.282376
	FBC	2366	-0.000349	0.115429	0.013324	
alcsummed2 = alcrisk*summed2;	Hosp	11475	0.000002	0.045614	0.002081	1.641470
	FBC	2366	-0.000010	0.035603	0.001268	
drugsummed2 = drugrisk*summed2;	Hosp	11475	0.000000	0.048450	0.002347	0.929016
	FBC	2366	-0.000004	0.050267	0.002527	
parsummed2 = parrisk2*summed2;	Hosp	11475	0.000134	0.231425	0.053557	2.526478
	FBC	2366	-0.000651	0.145597	0.021198	
mr19summed2 = mr19*summed2;	Hosp	11475	0.000052	0.148815	0.022146	2.467484
	FBC	2366	-0.000254	0.094737	0.008975	

Table B-4: Results from the log-linear analysis comparing the presence of a risk between groups with and without quintiles for FBC vs. Midwife-attended Hospital comparison using the *extended* propensity score model

	2-way log linear model Group*Risk		3-way log linear model Group*Risk*Quintile	
Group*Risk	Chi-Square	Pr>ChiSq	Chi-Square	Pr>ChiSq
<19 years old	11.99	0.0005	3.22	0.07
>35 years old	37.96	<.0001	0.11	0.74
<12 years education	15.03	0.0001	0.14	0.71
African American risk marker	18.41	<.0001	1.44	0.23
Unmarried	22.25	<.0001	1.98	0.16
>20 weeks (or 5 months) gestation at first prenatal care visit	8.31	0.0039	0.26	0.61
Cigarette smoking	3.04	0.0813	3.3	0.07
Drinking alcohol	0.09	0.7652	0.64	0.42
Use of drugs	1.22	0.2696	0.52	0.47
First pregnancy - parity risk	21.97	<.0001	0.39	0.53
Prev infant 4000+ grams	23.19	<.0001	0.58	0.45
Prev small for gestation	2.03	0.1538	0.15	0.70
Prev preterm	7.43	0.0064	0.01	0.93
Greater than 3 previous terminations - (miscarriage, abortion, still birth)	3.67	0.0555	2.4	0.12
Sum of medical risks	83.48	<.0001	0.1	0.75
Other	64.27	<.0001	0.44	0.51
youngedu = agrisk_y*edurisk;	11.05	0.0009	0.79	0.38
oldedu = agrisk_o*edurisk;	0.03	0.8572	0.52	0.47
youngrace = agrisk_y*racerisk;	15.34	<.0001	0.37	0.54
oldrace = agrisk_o*racerisk;	0.71	0.3998	0.26	0.61
raceedu = racerisk*edurisk;	6.22	0.0126	2.04	0.15
youngmarr = agrisk_y*nmarrrisk;	36.65	<.0001	2.32	0.13
oldmarr = agrisk_o*nmarrrisk;	5.9	0.0151	0.98	0.32
edumarr = edurisk*nmarrrisk;	45.06	<.0001	2.39	0.12
racemarr = racerisk*nmarrrisk;	46.41	<.0001	0.99	0.32
youngpn = agrisk_y*pncare_risk;	3.97	0.0463	2.97	0.08
oldpn = agrisk_o*pncare_risk;	n/a		n/a	
edupn = edurisk*pncare_risk;	0.8	0.3713	0.13	0.71
racepn = racerisk*pncare_risk;	1.38	0.2406	0.06	0.80
marpn = pncare_risk*nmarrrisk;	13.52	0.0002	0.17	0.68
youngcig = agrisk_y*cigrisk;	0	0.9689	0.63	0.43
oldcig = agrisk_o*cigrisk;	7.82	0.0052	0.25	0.61
racecig = racerisk*cigrisk;	5.12	0.0236	0.72	0.40
educig =edurisk*cigrisk;	1.9	0.1681	5.7*	0.02
cigmar =cigrisk*nmarrrisk;	1.29	0.2553	2.12	0.15
cigpn =cigrisk*pncare_risk;	7.7	0.0055	1.19	0.27
alcold = agrisk_o*alcrisk;	0.04	0.8432	0.15	0.70
alcrace = racerisk*alcrisk;	1.56	0.2123	0.04	0.84
alcmar = alcrisk*nmarrrisk;	0.2	0.6537	2.18	0.14
alcpn = alcrisk*pncare_risk;	n/a		n/a	
alccig = alcrisk*cigrisk;	0.25	0.6139	0.14	0.71
drugyoung = drugrisk*agrisk_y;	0.01	0.9434	1.19	0.28
drugedu = drugrisk*edurisk;	2.45	0.1173	0.56	0.46
drugrace = drugrisk*racerisk;	0.8	0.3709	0.01	0.91

drugmar = drugrisk*nmarrrisk;	0.2	0.6528	0.52	0.47
drugpn = drugrisk*pncare_risk;	0	0.9581	2.39	0.12
drugcig = drugrisk*cigrisk;	0.12	0.7252	0	0.95
drugalc = drugrisk*alcrisk;	0.1	0.7475	0.87	0.35
youngparr = agrisk_y*parrisk2;	4.16	0.0414	2.87	0.09
oldparr = agrisk_o*parrisk2;	4.47	0.0345	0.46	0.50
eduparr = edurisk*parrisk2;	3.27	0.0704	0	0.95
raceparr = racerisk*parrisk2;	0.72	0.3974	0.02	0.88
marparr = nmarrrisk*parrisk2;	21.1	<.0001	0.7	0.40
cigparr = cigrisk*parrisk2;	0.32	0.569	3.94*	0.05
drugparr = drugrisk*parrisk2;	4.14	0.0419	0.73	0.39
racemr19 = racerisk*mr19;	22.72	<.0001	0.52	0.47
marrmr19 = nmarrrisk*mr19;	31.9	<.0001	1.76	0.18
pnmr19 = pncare_risk*mr19;	3.28	0.0701	0.36	0.55
parmr19 = parrisk2*mr19;	9.12	0.0025	0.11	0.74
drugmr19 = drugrisk*mr19;	0.06	0.8134	0.49	0.48
youngsummed2 = agrisk_y*summed2;	23.21	<.0001	0.83	0.36
edusummed2 = edurisk*summed2;	49.31	<.0001	0.15	0.70
racesummed2 = racerisk*summed2;	12.4	0.0004	0.02	0.89
marsummed2 = nmarrrisk*summed2;	48.49	<.0001	0.33	0.57
cigsummed2 = cigrisk*summed2;	24.09	<.0001	1.05	0.30
alcsummed2 = alcrisk*summed2;	0.74	0.3905	0.15	0.70
drugsummed2 = drugrisk*summed2;	0.01	0.9358	0.15	0.70
parsummed2 = parrisk2*summed2;	50.42	<.0001	0.24	0.62
mr19summed2 = mr19*summed2;	15.06	0.0001	0.02	0.89

* $p<0.05$

Table B-5: Outcomes for the FBC-Midwife analysis controlled for the extended propensity score model

Outcomes[b]	-2 Log Likelihood		Likelihood Ratio Test				
	Intercept only model (null model deviance)	Intercept plus covariates model (full model deviance)	Chi-Square[a]	Pseudo R-Square	Max-rescaled R-Square	Adjusted Odds Ratio for group	95% Confidence Interval
LBW (<2500)	3900.644	3868.204	32.4401	0.0023	0.0095	0.74*	(0.55-0.99)
Small for gestational age	7513.561	7497.417	16.1446	0.0012	0.0028	0.80*	(0.67-0.96)
Large for gestational age	8394.779	8315.052	79.7276	0.0057	0.0126	1.45***	(1.26-1.67)
Preterm birth (<37)	5429.838	5386.188	43.6501	0.0031	0.0097	0.61***	(0.48-0.79)
Apgar score, <7 at 1	6302.007	6292.936	9.0708	0.0007	0.0018	1.07	(0.89-1.29)
Apgar score, <7 at 5	1105.85	1104.771	1.0789	0.0001	0.001	1.08	(0.62-1.87)
Fetal distress	7367.873	7223.361	144.5117	0.0104	0.0252	0.31***	(0.24-0.41)
Incidence of abnormal conditions of the infant	7016.006	6876.786	139.2206	0.01	0.0252	1.01	(0.84-1.22)
Incidence of maternal complications during labor	14079.028	13988.937	90.0906	0.0065	0.0102	0.85**	(0.76-0.96)

*p<0.05; ** p<0.01; *** p<0.001

[a] All with 2 degrees of freedom

[b] The proportion of total cases dropped from each logistic regression model because of missing information is approximately 0.019.

REFERENCES

Adamson, G. D., & Gare, D. J. (1980). Home or hospital births? *Journal of the American Medical Association, 243*, 1732-1736.

Albers, L. L., & Katz, V. L. (1991). Birth settings for low-risk pregnancies. An analysis of the current literature. *Journal of Nurse-Midwifery, 36*(4), 215-220.

Alexander, G.R., Himes, J.H., Kaufman, R.B., Mor J., & Kogan, M. (1996). A United States national reference for fetal growth. *Obstetrics & Gynecology, 87*(2), 163-168.

Alexy, B., Nichols, B., Heverly, M. A., & Garzon, L. (1997). Prenatal factors and birth outcomes in the public health service: a rural/urban comparison. *Research in Nursing & Health, 20*(1), 61-70.

Amato, J. C. (1977). Fetal monitoring in a community hospital: a statistical analysis. *Obstetrics & Gynecology, 50*, 269-274.

American Academy of Pediatrics and American College of Obstetricians and Gynecologists. (1988). *Guidelines for perinatal care* (2nd ed). Elk Grove Village (IL): AAP.

American Public Health Association. (1983). Position paper 7924: Alternatives in maternity care. APHA Public Policy Statements, 1948 to present, cumulative. *American Journal of Public Health, 73*(3), 331-334.

Austin, P. C., Mamdani, M. M., Stukel, T. A., Anderson, G. M., & Tu, J. V. (2005). The use of the propensity score for estimating treatment effects: administrative versus clinical data. *Statistics in Medicine, 24*, 1563-1578.

Baruffi, G., Dellinger, W. S., Stobino, D. M., Rudolph, A., Timmons, R. Y., & Ross, A. (1984). A study of pregnancy outcomes in a maternity center and a tertiary care hospital. *American Journal of Public Health, 74*(9), 973-978.

Bennetts, A. B. & Lubic, R. W. (1982). The freestanding birth centre. *Lancet, 1*(8268), 378-380.

Berkowitz, G. S. (1981). An epidemiologic study of preterm delivery. *American Journal of Epidemiology,113*(1), 81-92.

Bernstein, P., Harrison, E., & Merkatz, I. (2000). Preconception and prenatal care. In W. Cohen (Ed.), *Cherry and Merkatz's Complications of Pregnancy* (5th ed.). Philadelphia: Lippincott Williams and Wilkins.

Buehler, J., Prager, K., & Hogue, C. (2000). The role of linked birth and infant death certificates in maternal and child health epidemiology in the United States. *American Journal of Preventive Medicine, 19*(1 Suppl), 3-11.

Buescher, P. A., Smith, C., Holliday, J. L., & Levine, R. H. (1987). Source of prenatal care and infant birth weight: the case of a North Carolina county. *American Journal of Obstetrics and Gynecology, 156*(1), 204-210.

Buka, S. L., Brennan, R. T., Rich-Edwards, J. W., Raudenbush, S. W., & Earls, F. (2003). Neighborhood support and the birth weight of urban infants. *American Journal of Epidemiology, 157*(1), 1-8.

Cepeda, M. S., Boston, R., Farrar, J. T., & Strom, B. L. (2003). Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *American Journal of Epidemiology, 158*(3), 280-287.

Chomitz, V. R., Cheung, L., & Lieberman, E. (1995). The role of lifestyle in preventing low birth weight. *The Future of Children, 5*(1), 121-138.

Cochran, W. G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics, 24*(205-213).

Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Cohen, R. L. (1982). A comparative study of women choosing two different childbirth alternatives. *Birth, 9*(1), 13-19.

Cordray, D. S., & Davidson, H. A. (2004). *Kellogg Birth Center Project: Final evaluation report*. Nashville, TN: Vanderbilt University.

Cunningham, F. G., MacDonald, P. C., Grant, N. F., Leveno, K. J., Gilstrap, L. C., Hankins, G., et al. (Eds.). (1997). *Williams Obstetrics* (20th ed.). Stamford, Connecticut: Appleton & Lange.

Davidson, H. A. (2001). *The community freestanding birth center: An empirical synthesis linking theory with evidence.* Unpublished Master's thesis, Vanderbilt University, Nashville, TN.

Davidson, H. A., Cordray, D. S., & Johnson, L. (2002). *Choosing maternity care: Women's decision strategies help explain issues of access.* Paper presented at the American Public Health Association's 130th Annual Meeting and Exposition, Philadelphia, PA.

Davis-Floyd, R. (2001). The technocratic, humanistic, and holistic paradigms of childbirth. *International Journal of Gynecology and Obstetrics, 75*, S5-S23.

Dehejia, R. (2005). Practical propensity score matching: a reply to Smith and Todd. *Journal of Econometrics, 125*(1-2), 355-364.

Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics, 84*(1), 151-161.

DeJong, R. N., Shy, K. K., & Carr, K. C. (1981). An out-of-hospital birth center using university referral. *Obstetrics & Gynecology, 58*, 703-707.

Devitt, N. (1996). The transition from home to hospital birth in the United States, 1930-1960. In P. K. Wilson (Ed.), *Childbirth methods and folklore* (Vol. 3, pp. 199-210). New York: Garland Publishing.

DiGiuseppe, D. L., Aron, D. C., Ranbom, L., Harper, D. L., & Rosenthal, G. E. (2002). Reliability of birth certificate data: A multi-hospital comparison to medical records information. *Maternal and Child Health Journal, 6*(3), 169-179.

Dobie, S. A., Baldwin, L.M., Rosenblatt, R. A., Fordyce, M. A., Andrilla, C. H. A., & Hart, L. G. (1998). How well do birth certificates describe the pregnancies they report? The Washington State experience with low-risk pregnancies. *Maternal and Child Health Journal, 2*(3), 145-154.

Dunkel-Schetter, C., Gurung, R. A., Lobel, M., & Wadhwa, P. (2001). Stress processes in pregnancy and birth: Psychological, biological, and sociological influences. In A. Baum, T. Revenson & J. Singer (Eds.), *Handbook of Health Psychology* (pp. 495-518). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Eakins, P. S. (1984). The rise of the freestanding birth center: Principles and practice. *Women & Health, 9*(4).

Eakins, P. S., O'Reilly, W. B., May, L. J., & Hopkins, J. (1989). Obstetric outcomes at the Birth Place in Menlo Park: The first seven years. *Birth, 16*(3), 123-129.

Eden, R. D., Seifert, L. S., Winegar, A., & Spellacy, W. N. (1987). Perinatal characteristics of uncomplicated postdate pregnancies. *Obstetrics & Gynecology, 69*(296-299).

English, P. B., Kharrazi, M., Davies, S., Scalf, R., Waller, L., & Neutra, R. (2003). Changes in the spatial pattern of low birthweight in a southern California county: the role of individual and neighborhood level factors. *Social Science and Medicine, 56*, 2073-2088.

Faison, J. B., Pisani, B. J., Douglas, R. G., Cranch, G. S., & Lubic, R. W. (1979). The childbearing center: An alternative birth setting. *Obstetrics & Gynecology, 54*(4), 527-532.

Feldman, E., & Hurst, M. (1987). Outcomes and procedures in low risk birth: a comparison of hospital and birth center settings. *Birth, 14*(1), 18-24.

Feldman, P. J., Dunkel-Schetter, C., Sandman, C. A., & Wadhwa, P. D. (2000). Maternal social support predicts birth weight and fetal growth in human pregnancy. *Psychosomatic Medicine, 62*(5), 715-725.

Fleiss, J. L. (1992). General design issues in efficacy, equivalency and superiority trials. *Journal of Periodontal Research, 27*, 306-313.

Fullerton, J. (1982). The choice of in-hospital or alternative birth environment as related to the concept of control. *Journal of Nurse-Midwifery, 27*(2), 17-22.

Fullerton, J. T., & Severino, R. (1992). In-hospital care for low-risk childbirth. Comparison with results from the National Birth Center Study. *Journal of Nurse-Midwifery, 37*(5), 331-340.

Fullerton, J. T., & Wingard, D. (1990). Methodological problems in the assessment of nurse-midwifery practice. *Applied Nursing Research, 3*(4), 153-160.

Galotti, K. M., Pierce, B., Reimer, R. L., & Luckner, A. E. (2000). Midwife or doctor: A study of pregnant women making delivery decisions. *Journal of Midwifery & Women's Health, 45*(4), 320-329.

Geronimus, A.T. (1996). Black/White differences in the relationship of maternal age to birthweight: A population-based test of the weathering hypothesis. *Social Science and Medicine, 42* (4): 589-597.

Gould, J. B. (1999). Vital records for quality improvement. *Pediatrics, 103*(1 Suppl E), 278-290.

Gould, J. B., Danielsen, B., Korst, L. M., Phibbs, R., Chance, K., Main, E., et al. (2004). Cesarean delivery rates and neonatal morbidity in a low-risk population. *Obstetrics & Gynecology, 104*(1), 11-19.

Howell-White, S. (1999). *Birth Alternatives: How women select childbirth care.* Westport, Connecticut: Greenwood Press.

Hundley, V. A., Cruickshank, F. M., Lang, G. D., Glazener, C. M., Milne, J. M., Turner, M., et al. (1994). Midwife managed delivery unit: a randomised controlled comparison with consultant led care. *British Medical Journal, 309*(6966), 1400-1404.

Hundley, V. A., Milne, J. M., Glazener, C. M., & Mollison, J. (1997). Satisfaction and the three C's: continuity, choice and control. Women's views from a randomised controlled trial of midwife-led care. *British Journal of Obstetrics & Gynaecology, 104*(11), 1273-1280.

Iams, J. D., Goldenberg, R. L., Mercer, B. M., Moawad, A. H., Meis, P. J., Das, A. F., et al. (2001). The preterm prediction study: can low-risk women destined for spontaneous preterm birth be identified? *American Journal of Obstetrics & Gynecology, 184*(4), 652-655.

Ickovics, J. R., Ethier, K. A., Koenig, L. J., Wilson, T. E., Walter, E. B., & Fernandez, M. I. (2000). Infant birth weight among women with or at high risk for HIV infection: the impact of clinical, behavioral, psychosocial, and demographic factors. *Health Psychology, 19*(6), 515-523.

Institute of Medicine and National Research Council (1982). *Research issues in the assessment of birth settings.* Washington, D.C.: National Academy Press.

Institute of Medicine (1985). *Preventing low birthweight.* Washington, D.C.: National Academy Press.

Jackson, D. J., Lang, J. M., Dickinson, C. P., & Fullerton, J. T. (1994). Use of the nurse-midwifery clinical data set for classification of subjects in birth center research. *Journal of Nurse-Midwifery, 39*(4), 197-213.

Jackson, D. J., Lang, J. M., Swartz, W. H., Ganiats, T. G., Fullerton, J., Ecker, J., et al. (2003). Outcomes, safety, and resource utilization in a collaborative care birth center program compared with traditional physician-based perinatal care. *American Journal of Public Health, 93*(6), 999-1006.

James, S. A. (1993). Racial and ethnic differences in infant mortality and low birth weight: A psychosocial critique. *Annals of Epidemiology, 3*(2), 130-136.

Kennedy, H. P. (2000). A model of exemplary midwifery practice: Results of a Delphi study. *Journal of Midwifery & Women's Health, 45*(1), 4-19.

Kennedy, H. P. (2002). The midwife as an "Instrument" of care. *American Journal of Public Health, 92*(11), 1759-1760.

Kramer, M. S. (1987). Determinants of low birth weight: methodological assessment and meta-analysis. *Bulletin of the World Health Organization, 65*(5), 663-737.

Kramer, M. S., Seguin, L., Lydon, J., & Goulet, L. (2000). Socio-economic disparities in pregnancy outcome: why do the poor fare so poorly. *Pediatric and Perinatal Epidemiology, 14*, 194-210.

Leavitt, J. W. (Ed.). (1999). *Women and Health in America* (2nd ed.). Madison, Wisconsin: The University of Wisconsin Press.

Leveno, K. J., Cunningham, F. G., & Nelson, S. (1986). A prospective comparison of selective and universal electronic fetal monitoring in 34,995 pregnancies. *New England Journal of Medicine, 315*, 615-619.

Lieberman, E., & Ryan, K. J. (1989). Birthday choices. *New England Journal of Medicine, 321*, 1824-1825.

Lieberman, E., Ryan, K. J., Monson, R. R., & Schoenbaum, S. C. (1987). Risk factors accounting for racial differences in the rate of premature birth. *New England Journal of Medicine, 317*(12), 743-748.

Little, R. J. & Rubin, D. B. (2000). Causal effects in clinical and epidemiological studies via potential outcomes: Concepts and analytical approaches. *Annual Review of Public Health, 21*, 121-145.

Lobel, M., Dunkel-Schetter, C., & Scrimshaw, S. C. (1992). Prenatal maternal stress and prematurity: a prospective study of socioeconomically disadvantaged women. *Health Psychology, 11*(1), 32-40.

Lubic, R. W. & Ernst, E. K. M. (1978). The childbearing center: An alternative to conventional care. *Nursing Outlook,* 754-760.

Maine, D. (1991). *Safe motherhood programs: Options and issues*. New York, NY: Columbia University Press.

Martin, J., Hamilton, B. E., Sutton, P., Ventura, S. J., Menacker, F., & Munson, M. L. (2003). Births: Final data for 2002. *National Vital Statistics Reports, 52*(10).

Mathews, J. J., & Zadak, K. (1991). The alternative birth movement in the United States: History and current status. *Women & Health, 17*(1), 39-56.

McAnarney, & Stevens-Simon, C. (1990). Maternal psychological stress/depression and low birth weight. Is there a relationship? *American Journal of Diseases in Children, 144*, 789-792.

McCallum, W. T. (1979). The Maternity Center at El Paso. *Birth and the Family Journal, 6*(4), 259-266.

Michalopoulos, C., Bloom, H. S., & Hill, C. J. (2004). Can propensity-score methods match the findings from a random assignment evaluation of mandatory welfare-to-work programs? *Review of Economics and Statistics, 86*(1), 156-179.

Michielutte, R., Ernest, J. M., Moore, M. L., Meis, P. J., Sharp, P. C., Wells, H. B., et al. (1992). A comparison of risk assessment models for term and preterm low birthweight. *Preventive Medicine, 21*, 98-109.

Motulsky, H. (1995). *Intuitive biostatistics*. New York: Oxford University Press.

Nagelkerke, N. J. D. (1991). A note on the general definition of the coefficient of determination. *Biometrika, 78*, 691-692.

National Association of Childbearing Centers (1999). *How to start a birth center: Exploring innovation in maternity care.* Perkiomenville, PA: NACC.

Norbeck, J. S. & Tilden, V. P. (1983). Life stress, social support, and emotional disequilibrium in complications of pregnancy: A prospective multivariate study. *Journal of Health and Social Behavior, 24*, 30-46.

Oakley, D., Murray, M. E., Murtland, T., Hayashi, R., Andersen, H. F., Mayes, F., et al. (1996). Comparisons of outcomes of maternity care by obstetricians and certified nurse-midwives. *Obstetrics and Gynecology, 88*(5), 823-829.

Oakley, D., Murtland, T., Mayes, F., Hayashi, R., Petersen, B. A., Rorie, C., et al. (1995). Comparisons of certified nurse-midwives and obstetricians. *Journal of Nurse-Midwifery, 40*(5), 399-409.

Olsen, O. & Jewell, M. D. (2000). Home versus hospital birth. *Cochrane Database of Systematic Reviews*(2), CD000352.

Paarlberg, K. M., Vingerhoets, A. J. J. M., Passchier, J., Dekker, G. A., & Van Geijn, H. P. (1995). Psychosocial factors and pregnancy outcome: A review with emphasis on methodological issues. *Journal of Psychosomatic Research, 39*(5), 563-595.

Paine, L.L., Johnson, T.R., Lang, J.M., Gagnon, D., Declercq, E.R., DeJoseph, J., et al. (2000). A comparison of visits and practices of nurse-midwives and obstetrician-gynecologists in ambulatory care settings. *Journal of Midwifery & Women's Health, 45*(1), 37-44.

Paneth, N. S. (1995). The problem of low birth weight. *The Future of Children, 5*(1).

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology, 49*(12), 1373-1379.

Piper, J. M., Mitchel, E. F., Snowden, M., Hall, C., Adams, M., & Taylor, P. (1993). Validation of 1989 Tennessee birth certificates using maternal and newborn hospital records. *American Journal of Epidemiology, 137*(7), 758-768.

Raisler, J. (2000). Midwifery care research: What questions are being asked? What lessons have been learned? *Journal of Midwifery & Women's Health, 45*(1), 20-36.

Rauh, V. A., Andrews, H. F., & Garfinkel, R. S. (2001). The contribution of maternal age to racial disparities in birthweight: a multilevel perspective.[comment]. *American Journal of Public Health, 91*(11), 1815-1824.

Reinke, C. (1982). Outcomes of the first 537 births at the Birthplace in Seattle. *Birth, 9*(4), 231-238.

Rini, C. K., Dunkel-Schetter, C., Wadhwa, P. D., & Sandman, C. A. (1999). Psychological adaptation and birth outcomes: the role of personal resources, stress, and sociocultural context in pregnancy. *Health Psychology, 18*(4), 333-345.

Rooks, J., Weatherby, N., & Ernst, E. K. M. (1992a). The National Birth Center Study: Part 1 - Methodology and prenatal care and referrals. *Journal of Nurse-Midwifery, 37*(4), 222-253.

Rooks, J., Weatherby, N., & Ernst, E. K. M. (1992b). The National Birth Center Study: Part 2 - Intrapartum and immediate postpartum and neonatal care. *Journal of Nurse-Midwifery, 37*(5), 301-330.

Rooks, J., Weatherby, N., & Ernst, E. K. M. (1992c). The National Birth Center Study: Part 3 - Intrapartum and immediate postpartum and neonatal complications and transfers, postpartum and neonatal care, outcomes, and client satisfaction. *Journal of Nurse-Midwifery, 37*(6), 361-397.

Rooks, J. P. (1997). *Midwifery and Childbirth in America*. Philadelphia: Temple University Press.

Rooks, J. P., Weatherby, N. L., Ernst, E. K. M., Stapleton, S., Rosen, D., & Rosenfield, A. (1989). Outcomes of care in birth centers. The National Birth Center Study. *New England Journal of Medicine, 321*(26), 1804-1811.

Rosenbaum, P. R. (2002). *Observational Studies* (2nd ed.). New York: Springer-Verlag.

Rosenbaum, P. R. & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association, 79*(387), 516-524.

Rosenblatt, R. A., Dobie, S. A., Hart, L. G., Schneeweiss, R., Gould, D., Raine, T. R., et al. (1997). Interspecialty differences in the obstetric care of low-risk women. *American Journal of Public Health, 87*(3), 344-351.

Rubin, D. B. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine, 127*(8S), 757-763.

Rubin, D.B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology, 2*, 169-188.

Rubin, D. B. & Thomas, N. (1996). Matching using estimated propensity scores: Relating theory to practice. *Biometrics, 52*(1), 249-264.

Scott, J.R., Gibbs, R.S., Karlan, B.Y., & Haney, A.F. (2003). *Danforth's Obstetrics & Gynecology*. (9th ed) Philadelphia: Lippicott Williams & Wilkins.

Scupholme, A., & Kamons, A. S. (1987). Are outcomes compromised when mothers are assigned to birth centers for care? *Journal of Nurse-Midwifery, 32*(4), 211-215.

Scupholme, A., McLeod, A. G. W., & Robertson, E. G. (1986). A birth center affiliated with the tertiary care center: Comparisons of outcomes. *Obstetrics & Gynecology, 67*(4), 598-603.

Seshadri, L., & Mukherjee, B. (2005). A predictive model for cesarean section in low risk pregnancies. *International Journal of Gynaecology & Obstetrics, 89*, 94-98.

Sheehan, T. J. (1998). Stress and low birth weight: A structural modeling approach using real life stressors. *Social Science and Medicine, 47*(10), 1503-1512.

Shiono, P. H., & Behrman, R. E. (1995). Low birth weight: Analysis and recommendations. *The Future of Children, 5*(1), 4-18.

Shiono, P. H., Rauh, V. A., Park, M., Lederman, S. A., & Zuskar, D. (1997). Ethnic differences in birthweight: the role of lifestyle and other factors. *American Journal of Public Health, 87*(5), 787-793.

Stanton, A. L., Lobel, M., Sears, S., & DeLuca, R. S. (2002). Psychosocial aspects of selected issues in women's reproductive health: current status and future directions. *Journal of Consulting and Clinical Psychology, 70*(3), 751-770.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn and Bacon.

Taylor, S. E., Repetti, R. L., & Seeman, T. (1997). Health psychology: What is an unhealthy environment and how does it get under the skin? *Annual Review of Psychology, 48*, 411-447.

Turnbull, D., Holmes, A., Shields, N., Cheyne, H., Twaddle, S., Gilmour, W. H., et al. (1996). Randomised, controlled trial of efficacy of midwife-managed care. *The Lancet, 348*, 213-218.

Turner, R. J., Grindstaff, C. F., & Phillips, N. (1990). Social support and outcomes in teenage pregnancy. *Journal of Health and Social Behavior, 31*, 43-57.

van Der Hulst, L. A., van Teijlingen, E. R., Bonsel, G. J., Eskes, M., & Bleker, O. P. (2004). Does a pregnant woman's intended place of birth influence her attitudes toward and occurrence of obstetric interventions? *Birth, 31*(1), 28-33.

Varney, H., Kriebs, J. M., & Gegor, C. L. (Eds.). (2004). *Varney's Midwifery* (4th ed.). Sudbury, MA: Joans and Bartlett Publishers.

Ventura, S. J., Hamilton, B. E., Mathews, T. J., & Chandra, A. (2003). Trends and variations in smoking during pregnancy and low birth weight: Evidence from the birth certificate, 1990-2000. *Pediatrics, 111*(5 Part 2), 1176-1180.

Villar, J., Ba'aqeel, H., Piaggio, G., Lumbiganon, P., Miguel Belizan, J., Farnot, U., et al. (2001). WHO antenatal care randomised trial for the evaluation of a new model of routine antenatal care. *Lancet, 357*(9268), 1551-1564.

Wagner, M. (2001). Fish can't see water: The need to humanize birth. *International Journal of Gynaecology & Obstetrics, 75*, S25-S37.

Waldenstrom, U., Nilsson, C., & Winbladh, B. (1997). The Stockholm Birth Center Trial: Maternal and infant outcomes. *British Journal of Obstetrics & Gynaecology, 104*, 410-418.

Walsh, D., & Downe, S. M. (2004). Outcomes of free-standing, midwife-led birth centers: A structured review. *Birth, 31*(3), 222-229.

Wiener, J.B . (1998). Managing the iatrogenic risks of risk management. *Risk: Heath, Safety & Environments, 39*, 39-82.

Winkelmayer, W. C., & Kurth, T. (2004). Propensity scores: help or hype? *Nephrology Dialysis Transplantation, 19*, 1671-1673.

Winship, C., & Morgan, S. L. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology, 25*, 659-706.

Wise, P. H. (2003). The anatomy of a disparity in infant mortality. *Annual Review of Public Health, 24*, 341-362.

Yali, A. M., & Lobel, M. (1999). Coping and distress in pregnancy: An investigation of medically high risk women. *Journal of Psychosomatic Obstetrics and Gynecology, 20*, 39-52.

Zimmer-Gembeck, M. J., & Helfand, M. (1996). Low birthweight in a public prenatal care program: Behavioral and psychosocial risk factors and psychosocial intervention. *Social Science and Medicine, 43*(2), 187-197.