2016

Hospital Profiling of the Cesarean Delivery Procedure for the State of Georgia, 2012

Denise Frances Giles
Walden University

Walden University

College of Health Sciences

This is to certify that the doctoral dissertation by

Denise F. Giles

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Vibha Kumar, Committee Chairperson, Public Health Faculty
Dr. Xianbin Li, Committee Member, Public Health Faculty
Dr. Susan Nyanzi, University Reviewer, Public Health Faculty

Chief Academic Officer
Eric Riedel, Ph.D.

Walden University
2016

Abstract

Hospital Profiling of the Cesarean Delivery Procedure for the State of Georgia, 2012

by

Denise F. Giles

MPH, University of Alabama-Birmingham, 1989

BS, Creighton University, 1985

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Public Health

Walden University

February 2016

Abstract

Approximately 35.1% of live births for the state of Georgia were delivered by the cesarean delivery procedure with significant variation among hospitals. The purpose of this research was to develop a population-based hospital profiling methodology for study of the cesarean delivery procedure. This was a retrospective, observational design, using a 2012 linked dataset that included maternity deliveries from all nonfederal hospitals. The research was guided by Robson 10 Group Classification System, propensity score methodologies, and ethical precepts, for the development of hospital profiles and the study of variations in the cesarean delivery procedure. Key research questions aimed to determine whether hospital profiling methodologies differed according to risk adjustment methods and statistical techniques. Propensity score matching with stratification methods aimed to determine whether there were differences in patient treatment effects on the cesarean delivery outcome. Findings suggested there was a significant difference in hospital ranks and model effects according to the statistical technique and the risk adjustment methods applied. Propensity score matching with stratification demonstrated an increased risk of the cesarean delivery procedure across strata, with the majority of high risk patients situated in the 90th percentile ranges and questionable utilization practice among other strata. Applying profiling methodologies at the facility and population level could advance statewide quality improvement programs for the timely reduction in the variation of inappropriate utilization of the cesarean delivery procedure.

Hospital Profiling of the Cesarean Delivery Procedure for the State of Georgia, 2012

by

Denise F. Giles

MPH, University of Alabama-Birmingham, 1989

BS, Creighton University, 1985

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Public Health

Walden University

February 2016

Dedication

In gratitude — for my family and their unconditional support. For the many friends and colleagues and their encouraging words. For the Georgia Department of Public Health, Office of Health Indicators for Planning, and advancing novel public health practice and research approaches from which to learn from. For PhD dissertation committee members, their scientific review, along with, Walden University for its academic commitment to social change, and without the awarded Commitment to Social Change Scholarship, would not have achieved this terminal degree. For the Centers for Disease Control and Prevention in supporting my PhD training plan while working as a U.S. civil servant assigned to its global mission, in parallel. May God be the glory — in humble responsibility, may I remain steadfast to the path given.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction to the Study

The purpose of this study was to examine variations in hospital profiling methodologies for the cesarean delivery procedure using comparative risk adjustment models and statistical techniques. Propensity score matching with stratification was applied to the observational design for an assessment of variation in risk of the cesarean delivery procedure.

Background

Hospital profiling—or more commonly known as report cards, performance reports, or consumer reports—are aimed to describe variations in health care utilization or health related outcomes (Normand & Shahian, 2007). Administrative data sources were used to identify provider practice patterns for the benchmark comparison to accepted standards of care or comparison with other providers (Ash et al., 2012). Provider profiling introduced statistical methodologies for describing variation in health care practice and assisted in determining whether patterns appropriately met expected cost, utilization, and quality standards (Shahian et al., 2011).

When provider profiling methods emerged in the 1990s, it was in response to a U.S. health care system concerned with quality of care and cost containment, managed care insurance environment (Shahian et al., 2011). These approaches have remained relevant to U.S. health care reform, given mandates to improve quality of health care, benchmark performance, and consumer choice (Patient Protection & Affordable Care Act, 2010). Provider profiling tools were treated skeptically due to their questionable reliability (Hofer et al., 1999; Werner & Asch, 2005), while others considered them part of due diligence (Christianson, Volmar, Alexander, & Scanlon, 2010) and an important methodology (Krumholz et al., 2005). The routine use of profiling by health care organizations at the state (Racz & Sedransk, 2010; Shahian, Torchiana, Shemin, Rawn, & Normand, 2005) and federal levels (Orzag, 2008) had shared momentum. More compellingly, profiling methods were extended by social media, often in the hands of consumers, who based their evaluations on their own experiences (RateMDs.com, 2012). Provider profiling tools have advanced in their methodologies since they were first introduced in the 1990s. Its application to a different time, place, and health indicator was revisited in the statewide study of the cesarean delivery procedure for the state of Georgia.

Over the past decade, the U.S. cesarean delivery rate increased by approximately 69% (AHRQ, 2015). It is unknown if this shift represented evidence based practice or inappropriate variation due to patient, physician, hospital, or health systems determinants (Chaillet & Dumont, 2007; Joseph et al., 2003; Lin & Xirasagar, 2004). Over the last decades, the cesarean delivery procedure continued to receive attention, as noted by the generous production of quality improvement guidelines by U.S. federal and nonfederal organizations, with intentions of reducing inappropriate utilization (Main, 2009).

Standard profiling methods ranked providers for their comparison and examination of outliers (Ash et al., 2012). A fair comparison of hospital provider ranks benefitted from reliable approaches capable of describing inter and intraprovider effects on cesarean delivery utilization (Coonrod, Drachman, Hobson, & Manriquez, 2008; Fantini et al., 2006). Profiling methods lent themselves to critique, because the natural question when reviewing hospital ranks was not solely if they were accurate or different, but also, why did they differ.

Problem Statement

Among the 3.9 million U.S. births reported in 2014, approximately 32.7% were delivered by the cesarean section procedure (Martin, Hamilton, & Osterman, 2014). This represented a 69% increase when compared to a 1990 cesarean delivery rate of 22.7% (Martin et al., 2014). Cesarean delivery was the most common U.S. surgical procedure, representing 4.7% of total aggregated hospitalization costs or $7.7 billion per year (Elixhauser & Andrews, 2010). The mean cost per hospital cesarean delivery varied according to procedures with complications ($23,923) or without complications ($17,889) (AHRQ, 2015). The cost of a vaginal procedure also varied according to hospital deliveries with complications ($13,749), hospital deliveries without complications ($10,657), and birth center vaginal deliveries ($2,277) (AHRQ, 2015; AABC, 2010). The majority of U.S. deliveries were vaginal deliveries with no complications (57%) followed by a cesarean with no complications (22%), cesarean with complications (12%), and vaginal with complications (9%) (AHRQ, 2015). The cesarean procedure was preferred to a vaginal delivery if clinical indications threatened the health of the mother, infant, or both. Common clinical indications known to increase the likelihood of a cesarean delivery included failure to progress during labor, fetal distress, or breech presentation (Zhang, Troendle, Reddy, Laughon, Branch, Burkman, & et al., 2010).

National cesarean delivery rates may not have peaked yet and could mirror, or surpass, utilization practices of China (41%), Mexico (39%), Brazil (37%) or Italy (36%) (Betran et al., 2007). U.S. hospital trends demonstrated significant variation in cesarean delivery rates with a proportion of facility procedures exceeding 50% to a high of 69% (Kozhimannil, Law, & Virnig, 2013). In contrast, the U.S. cesarean delivery rate exceeded utilization practices of other industrialized nations including the Netherlands (14%), Ireland (23%), and Germany (23%) (Betran et al., 2007).

U.S. cesarean delivery rates were known to vary by hospital and geographical place. Differences were attributed to patient obstetric risk factors, medical malpractice norms, and contextual attributes associated with regional health care supply or socioeconomic status (Baicker, Buckles, & Chandra, 2006; Kozhimannil et al., 2013). Up to 40% of cesarean delivery utilization were attributed to other unknown determinants (Baicker et al., 2006) affecting the health services model. A proportion of cesarean deliveries represented an overuse of clinically unwarranted procedures resulting in an increase of inappropriate healthcare expenditures, and in some cases, a risk to the mother, and, or her infant (Kabir et al., 2004; Queenan, 2011; Scott, 2011; Srinivas, Fager, & Lorch, 2010).

Absent of U.S. health care reform law, financial incentives, or collective responsibility, there was limited initiative for significantly curbing unwarranted utilization of health care services. Rather, hospitals, physicians, and patients contributed to inappropriate care through a number of causes: (a) payment schemas that rewarded volume versus quality; (b) increased utilization in reproductive health technologies; (c) direct to consumer health care marketing, influencing consumer choice; and (d) malpractice laws, compelling defensive medical practice (Emanuel & Fuchs, 2008). Reinforcement of ethical precepts (Clancy, 2011; John Paul II, 1991), and reducing preventable complications and process inefficiencies, was part of the traditional (Leape, 1992) and the new call (Swensen et al., 2011) for greater provider responsibility.

State level strategies have addressed unwarranted cesarean delivery utilization for the appropriate reduction in surgical procedures. Beginning in 2009, Washington's Medicaid program reduced the reimbursement amount for uncomplicated cesarean deliveries to that of complicated vaginal delivery procedures (Thompson, 2009). Texas Medicaid restricted cesarean deliveries among births occurring before 39 weeks gestational age when it was deemed medically unnecessary. Up to 40% of U.S. deliveries were insured by Medicaid (DHHS, 2011), and this proportion was expected to increase, given the expansion under U.S. health care reform. To prepare for this mandate, the U.S. Department of Health and Human Services (DHHS) identified cesarean delivery efficiencies by reducing nonmedically necessary preterm deliveries among Medicaid eligible mothers (DHHS, 2011).

More comprehensive state reforms aimed to reduce cesarean deliveries irrespective of insurance type. For example, Vermont targeted inappropriate utilization among low risk women, assured appropriate vaginal births after cesarean deliveries, reduced induction of labor less than 39 weeks of gestational age absent clinical indication, and released cesarean delivery data to individual health care practitioners for quality reviews (Vermont, Reimbursement for Avoidable Cesarean Sections, H.392, 2011). Colorado's Healthcare Affordability Act introduced hospital monitoring and evaluation of the cesarean delivery procedure through quality improvement and patient safety initiatives (Colorado Department of Health Care Policy & Financing, 2010). New York State introduced legislation to decrease physician or midwife insurance premiums should they participate in continuing education for improving patient decision making with the intention of reducing unnecessary cesarean deliveries (New York Public Law S5153B-2011, 2011).

At the national level, organizations, including the Joint Commission, the National Quality Forum, and the National Quality Measures Clearinghouse, adopted quality improvement guidelines for the cesarean delivery procedure (AHRQ, 2015; Main, 2009). At times, organizational strategies appeared fragmented, with noted duplications of effort, and only some coordination among differing organizations (AHRQ, 2015; Main 2009). All aligned with the National Quality Strategy and complemented President Obama's Affordable Health Care Act, a strategy that sought systematic improvement in the U.S. quality of health care, its indicators, and for a reduction in expenditures (DHHS, 2011).

Notably, the increase in U.S. cesarean delivery rates may have eluded direct attention because the mandated reporting of its indicators were population-based, aggregate measures systematically reported at the national and subnational level using U.S. live birth certificates (NCHS, 2001; NCHS, 2012). The applied epidemiologic concepts failed to systematically measure appropriately health services utilizations and their variations at the hospital level (Kozhimannil et al., 2013; Orzag, 2008). Cesarean delivery utilization was a process measure describing the volume of surgical procedures for hospital maternity deliveries (ACOG, 2015). Reducing variation in practice patterns required a deliberate measurement and reporting of its heterogeneous effects due to patient, provider, community, and health systems structural determinants.

There were notably few statewide hospital profiling methodologies for the systematic monitoring of cesarean delivery utilization rates (Coonrod et al., 2008). Hospital comparison reports were often published which described variations in total cesarean delivery rates, and absent risk adjustment, but were prone to bias or unfair comparisons (Illinois Department of Public of Health, 2012). More recently proposed quality improvement indicators acknowledged cesarean delivery was a heterogeneous concept that varied according to the characteristics of the pregnancy, previous obstetric record, course of labor and delivery, and gestational age (Bailit et al., 1999; Coonrod et al., 2008; Keeler et al., 1997). This concept aligned with Robson 10 Group Classification, a quality improvement (QI) improvement approach introduced at the international level (Robson, Scudamore, & Walsh, 1996) and endorsed by the World Health Organization (WHO) as an effective tool in reducing inappropriate cesarean section utilization (WHO, 2015). Monitoring the effective management of the vaginal delivery procedure and reducing inappropriate labor induction (Zhang et al., 2010), cesarean deliveries occurring less than 39 weeks gestational age (Tita et al., 2009), low risk primary cesarean deliveries (Coonrod et al., 2008; Main, Bloomfield, & Hunt, 2004), and breech presentation (ACOG, 2006), were central to decreasing unwarranted utilization effectively at the facility and population levels.

Purpose of the Study

The purpose of this study was to develop a statewide, hospital profiling methodology for the examination of cesarean delivery utilization using population-based, linked data from the state of Georgia. Robson's 10 Group Classification System (Robson et al., 1996) was used for developing hospital profiles at the state and facility levels. Cesarean delivery was the dependent variable and defined as the number of procedures per total number of live births. Robson 10 Group Classifications, as well as clinical and sociodemographic independent variables were used as risk adjusters to model variation in the cesarean delivery outcome. To advance provider profiling statistical techniques, this dissertation applied propensity score risk adjustment methods to the hospital profiling methodology. Multilevel statistical modeling was used to examine variation in cesarean delivery hospital ranks and risk adjustment effects. Findings from this research may be applied to health care reform, given the need for appropriate health care utilization and assuring quality of care.

Study Questions and Hypotheses

This research was guided by seven study questions and hypothesis tests for the comparative evaluation of the cesarean delivery outcome according to risk adjustment, hospital profiling, and propensity score methodologies.

1. RQ 1: Is there a difference in the mean cesarean delivery rate among the Robson 10 Groups?

$H_0$1: There is no difference in the mean cesarean delivery rate among the Robson 10 Groups.

H_a1: There is a difference in the mean cesarean delivery rate among the Robson 10 Groups.

2. RQ 2: Is there a difference in cesarean delivery risk adjustment effects when comparing logistic regression versus hierarchical generalized linear models?

H_02: There is no difference in cesarean delivery risk adjustment effects when comparing logistic regression versus hierarchical generalized linear models.

H_a2: There is a difference in cesarean delivery risk adjustment effects when comparing logistic regression versus hierarchical generalized linear models.

3. RQ 3: Is there a difference in cesarean delivery risk adjustment effects when comparing logistic regression propensity score versus hierarchical generalized linear propensity score models?

H_03: There is no difference in cesarean delivery risk adjustment effects when comparing logistic regression propensity score versus hierarchical generalized linear propensity score models.

H_a3: There is a difference cesarean delivery risk adjustment effects when comparing logistic regression propensity score versus hierarchical generalized linear propensity score models.

4. RQ 4: Is there a difference in cesarean delivery hospital ranks when comparing logistic regression versus hierarchical generalized linear models?

H_04: There is no difference in cesarean delivery hospital ranks when comparing logistic regression versus hierarchical generalized linear models.

H_a4: There is a difference in cesarean delivery hospital ranks when comparing logistic regression versus hierarchical generalized linear models.

5. RQ 5: Is there a difference in cesarean delivery hospital ranks when comparing logistic regression propensity score risk adjustment versus hierarchical generalized linear propensity score risk adjustment models?

H_05: There is no difference in cesarean delivery hospital ranks when comparing logistic regression propensity score risk adjustment versus hierarchical generalized linear propensity score risk adjustment models.

H_a5: There is a difference in cesarean delivery hospital ranks when comparing logistic regression propensity score risk adjustment versus hierarchical generalized linear propensity score risk adjustment models.

6. RQ 6: Is there a difference in cesarean delivery risk adjustment effects when comparing the propensity score matched sample versus the observational design sample?

H_06: There is no difference in cesarean delivery risk adjustment effects when comparing the propensity score matched sample versus the observational design sample?

H_a6: There is a difference in cesarean delivery risk adjustment effects when comparing the propensity score matched sample versus the observational design sample?

7. RQ 7: For the propensity score matched sample, is there a difference in cesarean delivery risk adjustment effects by stratum?

H_07: For the propensity score matched sample, there is no difference in cesarean delivery risk adjustment effects by stratum.

H_a7: For the propensity score matched sample, there is a difference in cesarean delivery risk adjustment effects by stratum.

Theoretical Frameworks

An interdisciplinary was approach used for the hospital profiling and propensity score research, borrowing strength from epidemiology and statistical theory. It was applied to a cesarean delivery classification system and guided by ethical precepts.

Robson 10 Group Classification System

Robson 10 Group Classification System (Robson et al., 1996) was used to develop a population-based hospital profiling methodology for the state of Georgia. At the international level, Robson's tool was first used at the hospital level for the prospective monitoring of cesarean deliveries and aided quality improvement practice (Robson et. al, 1996). The classification system organized pregnancy outcomes into 10 discrete categories for the review of delivery subgroups and associated rates (Table 1). By adapting Robson's framework, more recently proposed, U.S. population-based cesarean delivery (CS) indicators were harmonized into one system for routine monitoring and the standard comparison of cesarean deliveries at the state, national, or global levels (Betran et al., 2007; Brennan, Robson, Murphy, & O'Herlihy, 2009; Denk, Kruse, & Jain, 2006; Zhang et al., 2010). Robson 10 Group descriptive statistics and risk adjustment methods provided insights into advancing the study of the systematic variation in health outcomes and their inequities, as well as identifying solutions through health care reform.

Table 1

Description of Robson 10 Group Classification System

Robson Group	Description
Group 1	Nulliparous, single cephalic, ≥ 37 weeks, in spontaneous labor
Group 2	Nulliparous, single cephalic, ≥ 37 weeks, induced
Group 3	Multiparous (no previous CS), single cephalic, ≥ 37 weeks, in spontaneous labor
Group 4	Multiparous (no previous CS), single cephalic, ≥ 37 weeks, induced
Group 5	Previous uterine scar, single cephalic, ≥ 37 weeks
Group 6	All nulliparous breeches
Group 7	All multiparous breeches, with or without previous uterine scar
Group 8	All multiple pregnancies, with or without previous cesarean delivery
Group 9	All single pregnancies with abnormal lies, with or without uterine scar
Group 10	All single cephalic, ≤ 36 weeks, with or without previous cesarean delivery

Provider Profiling Theory

Provider profiling methods aimed to produce ranks for the fair comparison of patterns of practice. The approach was based in quality improvement theory, recognizing that health outcomes vary significantly according to provider structures and processes (Donabedian, 2002). When applied at the population level, information was used to understand variation in health outcomes and for the targeted intervention of reducing unwarranted practice, given its influence on health care costs and its importance in much needed reform. Model approaches typically applied multilevel statistical models, recognizing that organizations were characteristic of patients nested in hospitals, and variation existed within and between hospitals (Raudenbush & Byrk, 2002). Risk adjustment methods were used predominantly to model variation in patient populations and measured outcomes and controlled for confounding (Iezzoni, 1997).

Propensity Score Theory

The potential outcomes framework recognized two possible treatments for patient assignment in relation to an outcome, defined as treatment and control groups (Rosembaum & Rubin, 1983). For randomized controlled trials, treatment and control assignments were conducted a priori in an effort to reduce the likelihood of differences between subject assignments. Randomization produced an unbiased estimate of the average treatment effect, or the measure of moving an entire population from an untreated to a treated assignment (Austin, 2011). The observational design lacked randomization, resulting in systematic differences between the treatment and control groups. Propensity scores were used instead to estimate the average treatment effects for the observational design (Rosenbaum & Rubin, 1983).

For the cesarean delivery procedure, random assignment of an obstetrical patient to either a vaginal or cesarean delivery was deemed unethical. Absent of random assignment, observational studies employed varying study design and statistical approaches, with the aim of reducing threats to internal and external validity, and for the review of derived measures of association. A standard cesarean delivery risk adjustment model was not identified. Instead, differing patient, physician, provider, and community covariates were often used as risk adjustors to control for confounding. Cesarean delivery rates are known to vary significantly within or between hospitals given the type of patients entering into provider services (Bailit et al., 1999; Coonrod et al., 2008; Keeler et al., 1997; Leung et al., 1998). However, limited attention was given to understanding how patient level covariates were distributed across hospitals or health systems. This dissertation proposed propensity score risk adjustment methods for hospital profiling of the cesarean delivery indicator by estimating the propensity score for the log odds of receiving treatment. Additional post matching analyses were applied, including stratification based on the matched sample.

Propensity score methods balanced the data by matching cases (treatment) patients to controls on probabilities of receiving the treatment (Rosenbaum & Rubin, 1983). The probability of receiving the treatment was estimated through a propensity score which is a “conditional probability of assignment to a particular treatment group given a vector of observed covariates” (Rosenbaum & Rubin, 1983, p. 41). When applied to provider profiling, propensity scores aim to balance pretreatment variables among multiple providers, given observed patient level covariates. They reduced threats of selection bias, confounding, and biased effects of indicator measurement. Once matching was conducted, multivariate analyses were applied and comparable to randomized control trial methods. The multivariate analysis or multilevel models used the propensity scores as a sampling weight, where adjustments were made of observations based on the probabilities of inclusion in the sample. Regression models were applied to adjust and estimate average treatment effects by using a dichotomous explanatory variable indicating treatment conditions.

First, Do No Harm

Whether U.S. cesarean delivery utilization patterns had significant recourse, or if norms more likely valued a high technology interventionist strategy for labor and delivery (Denk et al., 2006), was yet to be determined at the start of this study. Although the technology existed, it too, can be inappropriately used to benefit tradeoffs, and in contradiction to the innateness of the human person and societal good. If the human person and social good is not considered, the individual, health care professionals, and health care systems can be considered to be merely elements of a production line, and medical decisions can be made based on what is most convenient or what is least likely to provoke lawsuits (John Paul II, 1991; Schifrin & Cohen, 2012). If defensive medicine or convenience of services is a new scientific standard for quality of care, then the new norm enables the selection into the cesarean delivery procedure as the most technological option (Burns, Geller, & Wholey, 1995; Schifrin & Cohen, 2012). Yet, it is expected for the majority of women to deliver a live born infant by vaginal delivery (Ye, Betran, Vela, Souza, & Zhang, 2014).

To advance hospital profiling methodologies, this research proposed scientific approaches balanced by ethical precepts (Clancy, 2011; John Paul II, 1991). In reviewing hospital profiles, and their risk adjusted rates, the intent was to compare their distribution accurately and detect outliers. Yet, when examined, seemingly normative rates may represent outliers instead. As science cannot fully intuit what is best for patients, ethical precepts were proposed to guide decision making. Subsidiarity decision making (Hamel & Nairn, 2011; John Paul II, 1991) may strengthen the role of personal responsibility, requiring a discussion between a physician and maternity patient regarding her own care, and thus the identification of quality improvement solutions were locally defined and relied upon. This precept assumed that the physician was the local authority (Hamel & Nairn, 2011) and that the medical decision making was technically competent, derived from compassion, and sought to avoid harm to the mother and her infant. It also assumed that health care organizations or governance structures had limited roles in the patient and physician dialogue, only serving to assist when appropriately needed, versus mandating or replacing this lowest level of autonomy (Hamel & Nairn, 2011). In complement, solidarity sought the common societal good while respecting local autonomy and its derived decision making between the patient and her physician (John Paul II, 1991; Sage, 2009). Unless these ethical decision making procedures are considered, high cesarean delivery rates were at risk of being accepted and normalized instead of seeking health systems solutions aimed at reducing inappropriate utilization for the better good of society.

Nature of the Study

This was a quantitative study for the hospital profiling of the cesarean delivery procedure in the state of Georgia. Comparative statistical and risk adjustment methods were applied to the observational design for the examination of variations in hospital ranks, model effects, and measures of association. Comparative statistical models were selected. The existing hospital profiling studies of the cesarean delivery outcome applied logistic regression methods; however, hierarchical generalized linear modeling were more accurate given the nested features of patients situated in hospitals (Raudenbush & Byrk, 2002). Comparative risk adjustment models were used in model development with the introduction of Robson Classification indicators versus more traditional approach of including clinical and sociodemographic characteristics alone. Propensity score matching with stratification was applied to the observational design as an extension to the hospital profiling methodology. When applied to the observational design, propensity score methods acted as an alternative to the randomized control trial by balancing baseline covariates of the treatment and control group for modeling model effects, hospital ranks, and measures of association.

Definitions

The dependent variable was the method of delivery and included the cesarean and vaginal procedure as response categories. The independent variables included Robson 10 Group Classification, clinical, sociodemographic, and propensity score characteristics. Other relevant terminologies were concerned with methodological approaches to the study design.

15-digit unique identification number (ID) — A 15-digit patient identifier defined as the first two letters of a patient's first name, first two letters of the last name, last two letters of the last name, patient date of birth, and patient sex (Giles, Austin, & Freymann, 2010).

Cesarean section (CS) —The birth of an infant via the abdominal route as a result of a blunt uterine incision (Berghella, 2007).

Clinical characteristics —Low-, medium-, and high-risk clinical conditions were modelled as independent variables and having an association with the cesarean delivery procedure (Bailit & Garrett, 2003).

Concatenated 14-digit unique identification number (ID)—A variable extraction method used to eliminate the first six and last alphanumeric characters from the 14-digit unique ID resulting in an 8-digit date of birth (Giles et al., 2010).

Concatenated 15-digit unique identification number (ID) — A variable extraction method used to eliminate the first six and last alphanumeric characters from the 15-digit unique ID resulting in an 8-digit date of birth (Giles et al., 2010).

Deterministic data linking—An exact matching methodology used to produce linked maternal records using standard data elements common to the 2012 Georgia hospital discharge summary and 2012 Georgia live birth files (Giles et al., 2010).

Hierarchical generalized linear models (HGLM)—Multilevel statistical techniques for nonlinear structural models (Raudenbusch & Bryk, 2002).

Hospital profiling—A statistical technique used to compare a provider's structure, processes of care, or outcomes, as a standard (Normand & Shahian, 2007).

Propensity score (PS)—The conditional probability of assignment to a treatment group given a vector of observed covariates (Rosenbaum & Rubin, 1983, p. 41).

Risk adjustment—A statistical technique used to adjust for variation in patient populations and measured outcomes (Iezzoni, 1997).

Robson 10 Group Classification—A clinical audit method used to classify cesarean and vaginal outcomes into 10 discreet groups according to characteristics of pregnancy, previous obstetric record, course of labor and delivery, and gestational age (Robson et al., 1996).

Sociodemographic characteristics (SES)—Age, and insurance type, and characteristics associated with the cesarean delivery outcome and change over time (Ash et al., 2012).

Assumptions

Hospital profiling and propensity score methods were guided by key assumptions but could not be demonstrated by the study findings. The first assumption was the number of hospitals, and their risks, represent accurate model estimates and ranks. This assumption was necessary because the study design was population-based and included all nonfederal hospitals for the state of Georgia but did not include federal hospitals or non-institutional births, or less than 5% of the sample. Second, risk adjustment and propensity score methods represented limited selection bias in modeling the cesarean delivery outcome because patient characteristics were measured and accounted for, measurable but not accounted for, or were difficult or impossible to measure (Ash et al., 2012). Lastly, an assumption was made that propensity score matching with stratification reduced bias in the estimation of treatment effects using the observational design (Rosenbaum & Rubin, 1983), recognizing that there were no identified simulation studies to guide the measurement of true effects.

Scope and Delimitations

This was a population-based hospital profiling study of the cesarean delivery procedure for the state of Georgia and included all live birth deliveries from nonfederal hospitals for the year 2012. The sampling frame was defined by a linked data file representing hospital discharge summary data for nonfederal maternal deliveries. Hospital profiling was used because of a higher than expected variation in the cesarean delivery outcome among Georgia hospitals, proposed risk adjustment, and statistical techniques for the reduction of inappropriate utilization, as well as to enable fair comparison of ranks. The study design aimed to improve internal validity by using a population-based linked dataset to ascertain risk adjustment variables that may not otherwise be systematically available through unlinked datasets. Findings from the hospital profiling methodology may be generalizable to the state of Georgia but questionable for other states, given variations in patient case mix, health care norms, and policies.

Propensity score matching with stratification was applied to the observational design as an extension to the hospital profiling methodology. The propensity score matched sample of cases (treatment) and controls allowed for the conditioning of observed effects of a nonrandomized design, given the difficulty of randomizing women to a cesarean or vaginal delivery. The matched sample's propensity score was based on data reduction techniques for one-to-one matching and characteristic of a quasi-randomized control trial design. Stratifying the propensity score matched sample allowed for the examination risk distribution across 10 subclasses. Propensity score matching was recognized as improving internal validity by balancing covariates, yet it was reliant upon appropriate variable selection and matching methods (Rosenbaum & Rubin, 1983). The higher the propensity scores by subgroup classification, the greater the risk for cesarean delivery. External validation of propensity score matching with stratification using simulation methods were required but lacking in the literature.

Limitations

Study limitations were concerned with potential selection bias of resulting hospital profiling and propensity score methodologies. To minimize this, data linking procedures were introduced to assure ascertainment of potential confounders which otherwise would not be available through unlinked sources. Comparative models were also introduced to understand potential biases in risk adjustment and statistical technique.

Significance

U.S. health care reform of reducing unwarranted utilization of the cesarean delivery procedure demanded solutions and at all levels of society (Clancy, 2011). This dissertation proposed a statewide, hospital profiling methodology and propensity score matching with stratification for the fair evaluation of hospitals and the cesarean delivery procedure. U.S. trends presented tighter regulations in order to reduce inappropriate cesarean delivery utilization among Medicaid insured patients and to obviate this demand for non-Medicaid and uninsured patients (DHHS, 2011; Thompson, 2009). Proposed public policies segmented risk pools according to insurance type versus patient medical risk and their systematic representation. However, a significant proportion of inappropriate utilization may also be due to non-Medicaid risk pools enabled by the purchasing power of patients, the provider practice, and the health systems they selected. When applied to studies of variation, the aforementioned strategy reinforced the need for unbiased approaches to assure the systematic examination of the quality of health systems. This structure of care was vulnerable to unnecessary harm. Health decisions were no longer the sole responsibility of the physician, but involved the interrelated relationships of patient autonomy, skilled health care providers, and in a framework of ethical health care policy.

Findings from this study may assist in identifying patient or hospital effects attributing to variation in cesarean delivery outcomes. Understanding geographic variation in health outcomes may introduce health care reform policy, guidelines, or intervention practices that are appropriate, equitable, and aim to reduce unnecessary health care expenditures for our nation.

Summary

Hospital profiling and propensity score matching with stratification were employed to advance quality improvement methodologies at the state and local level. In the comparison of cesarean delivery hospital ranks, hierarchical generalized linear modeling and risk adjustment models including propensity scores may present accurate methods when compared to the logistic regression methods predominantly used in maternal child health studies. Propensity score matching with stratification methods introduced a new approach on conditioning the cesarean delivery treatment and stratifying to examine risk according to stratum.

In addition to Chapter 1, Chapter 2 provides the literature review, Chapter 3 describes the research methodology, Chapter 4 reports on study findings, and Chapter 5 discusses the results and identified recommendations for future research and social change.

Chapter 2: Literature Review

Introduction

U.S. cesarean delivery rates were known to vary significantly among hospitals and a proportion of surgical procedures represented inappropriate utilization (Kozhimannil et al., 2013). Statewide hospital profiling and propensity score methods introduced strategies for reducing variation through quality improvement programs. The literature review referenced reproductive health, health services, the statistical methods for the appraisal of evidence, and guided research.

Literature Search

The literature research strategy reviewed electronic databases, dissertations, and websites. Electronic databases included CINAHL, Dissertations & Theses, Dissertation & Theses at Walden University, ERIC, Google Books, Google Scholar, ProQuest Central, PubMed/MEDLINE, SAGE Premier, Science Direct, Thoreau Multi-Database Search, OVID, Elsevier, Springer, Inform, and Cochrane Database of Systematic Reviews. Other organizational websites routinely searched included WHO, CDC, ACOG, and AHRQ. Key words used to search as concept or in combination with other terminologies included: *cesarean delivery, data linking, Robson 10 Group Classification, provider profiling, hospital profiling, propensity score, propensity score matching, propensity score stratification, risk adjustment, logistic regression, hierarchical generalized linear modeling, health care reform, quality improvement, maternal child health, obstetrics, and ethics.* No restrictions were placed on search terms according to publication date, geographical place, or whether the article was in English or another language. Real time alerts from research databases for selected terms were used to assure the information was as up to date as possible.

Population-based, Cesarean Delivery Indicators

A historical example in the review of the cesarean delivery procedure involved a letter to the *California State Journal of Medicine* by Spalding, (Spalding, 1910). Spalding reviewed findings from a 900 patient study where he described variation in rates of contracted pelvis ranging from 8% in clinical patients to 11% in private and hospital patients (Spalding, 1910). His elaboration further conveyed, “the idea that here in California, patients with contracted pelvis are a rarity is erroneous; the idea has gained ground because no one has made systematic examinations” (p. 50). In his later discussion of nine cases presented to the *Journal*, Spalding’s careful attention acknowledged that the cesarean delivery procedure occurred at rates higher than expected, inappropriate procedures were anticipated in the future, and skilled birth attendants were required if the operational procedure was to meet its intended purpose (Spalding, 1910).

By the 1980s, experts advocated for the safe reduction of U.S. cesarean delivery rates (Berwick, 1994), and others continued to advance these concerns today (Freeman et al, 2000; Queenan, 2011). Over 30 years ago, a number of quality improvement studies emerged with the purpose of safely reducing cesarean delivery rates (Dillon et al., 1992; Flamm, Berwick, & Kabcenell, 1998; Gregory, Hackmeyer, Gold, Johnson, & Platt, 1999; Kazandjian & Lied, 1998; Main, 1999; Myers & Gleicher, 1991). More recent U.S. quality improvement initiatives advanced novel approaches for reducing unwarranted variation in induction procedures, unplanned cesarean deliveries, and elective deliveries occurring before 39 weeks gestational age (Donovan et al., 2010; Fisch, English, Pedaline, Brooks, & Simhan, 2009; James & Savitz, 2011; Oshiro, Henry, Wilson, Branch, & Varner, 2009).

Quality improvement studies differed from more conventional epidemiologic observational designs because they applied clinical evidence to the healthcare environment for an assessment of its effectiveness in the management of practice patterns and effects on outcomes (Nicolay et al., 2012). Unlike randomized control trials, the generalizability of findings was questionable because evaluation methods were typically locally determined, highly heterogeneous, and used for multiple purpose (Farley & Battles, 2009). Their slow adoption was impeded by research strategies valuing hypothesis testing and generation of scientific evidence versus process improvement and experiential learning based on an established set of standards (Farley & Battles, 2009). Given a renewed emphasis on describing geographical variation in healthcare utilization, advanced quality improvement research methods were needed to advance U.S. healthcare reform intentions (Orzag, 2008).

At the national level, the U.S. was not complacent in monitoring cesarean delivery utilization (Martin et al., 2014). By 2000, the U.S. introduced two cesarean delivery indicators for the systematic monitoring of population-based rates using live birth certificates and reported at the national and subnational levels (NCHS, 2001). The nulliparous-vertex-singleton-cesarean delivery indicator (NTVS-CS), or a metric for low risk pregnancies, aimed to reduce this rate from 18% to 15%, and yet it escalated to 26.5% by 2007 instead (NCHS, 2001). The second indicator, a measure for repeat cesarean delivery, suffered similar misfortune, and instead of reaching its targeted goal of 63%, this metric escalated to 90.8% (NCHS, 2001). By 2012, the U.S. sustained these two national indicators for national monitoring with the goal of reducing the NTVS-CS rate to 23.5% and the repeat cesarean delivery rate to 81.9% (NCHS, 2012).

Even though the NVTS-CS indicator was a national mandate, limited attempts existed at validating its performance at the hospital (Zhang et al., 2010) or population level (Kahn, Berg, & Callaghan, 2009). Moreover, the scientific evidence supporting this specific indicator was also limited with the majority of cited studies based on informational sources derived around or before 2005 (Coonrod et al., 2008; Main et al., 1999; Main et al., 2004; Main et al., 2006).

With the American failure to meet national goals for the cesarean delivery objectives, there emerged a host of indicator guidelines sponsored by other nonfederal, federal, and local organizations (AHRQ, 2015; Main, 2009). These initiatives differed from the federal indicators because locally specific, quality improvement strategies were introduced with the aim of applying clinical evidence to the healthcare environment for an assessment of its effectiveness in the management of practice patterns and effects on outcomes (Nicolay et al., 2011). Hospital specific strategies aimed to inform cost, volume, and practice patterns of providers and determined whether the service, setting, and quality associated with health care received were appropriate or not (Lavis & Anderson, 1996). More recent indicator initiatives, nonsystematic and voluntary, supported provider autonomy with limited or no accountability in reporting to governmental entities. Instead, any indicator change in rates depended on inferences derived from national and subnational live birth certificate data because there was a lack of process improvement information at the population level (NCHS, 2012).

Although multiple indicator guidelines were emerging, attempts to describe cesarean delivery utilization conceptually according to health services quality improvement (Donabedian, 2002; Kesmodel & Jolving, 2011; Kilbourne, Fullerton, Dausey, Pincus, & Hermann, 2010) or sociobehavioral theory (Phillips, Morrison, Andersen, & Aday, 1998) were generally lacking. Instead, the focus remained on the introduction of new or strengthening existing cesarean delivery evidence based guidelines, which lacked guidance on quality improvement and its systematic implementation (ACOG, 2006; ACOG, 2007; ACOG, 2009; ACOG, 2010). Moreover, quality improvement approaches and cycles of learning for an understanding of their effects on outcomes were required. For example, multifaceted approaches including their continuous application through audit and feedback, peer review, peer leadership, or public accountability, were known to reduce cesarean delivery rates at the provider or health systems level (Chaillet & Dumont, 2007).

Relevant cesarean delivery utilization guidelines were often cited for improved medical decision making, included vaginal birth after cesarean delivery (ACOG, 2010), labor induction (ACOG, 2009), maternal request (ACOG, 2007), and breech presentation (ACOG, 2006). The majority of studies supporting cesarean delivery clinical guidelines and quality improvement indicators were based on retrospective observational designs with limited evidence generated from prospective quality improvement research (Bailit, Dooley, & Peacemen, 1999; Bailit & Garrett, 2003; Coonrod et al., 2008; Main et al., 2006). This divergence limited the development of the evidence base and an understanding of how quality improvement interventions improved healthcare and the appropriate utilization at the local or health systems level.

Less attention was given to creating frameworks for the effective integration of indicators into governance structures, their program evaluation, and understanding the systematic effect on health systems (Farley & Battles, 2009; Klazinga, Fischer, & Asbroek, 2011; Profit et al., 2010). Rather, existing approaches assumed providers and health systems had the political will, workforce capacity, and were equipped with the necessary quality improvement management systems to introduce indicators effectively into organizations for ongoing program evaluation and systematic change (Farley & Battles, 2009; Klazinga et al., 2011). Without ties to accreditation, national strategies, or policies, the systematic adoption of cesarean delivery indicators, and their meaningful use at the hospital level, was questionable, and so were attempts at reducing inappropriate utilization systematically and in a timely way (Farley & Battles, 2009; Klazinga et al., 2011; Profit et al., 2010).

Limited attention was also given to hospital based testing of indicator performance due to threats to validity and early adoption. Selection bias was known to threaten indicator development due to coding practices, the quality of data sources, and the abstraction methods used (Cheschier & Meints, 2009; Steinbush, Oostenbrink, Zuurbrier, & Schaepkens, 2007; Watkins et al., 2011). Selection bias was known to also affect risk adjusted outcomes due to statistical techniques, informational sources, or risk adjustment variable selection. Not as readily discussed was the effect of selection bias when indicators reported aggregated level rates versus identifying the appropriate unit of analysis based on multilevel effects of health services organization (Ash et al., 2012; Raudenbush & Byrk, 2002).

Administrative biases also existed based on the type of health care organization of hospitals from which rates were derived. Indicators suffered from misclassification bias due to the inability to distinguish between diagnoses present on admission or whether cesarean deliveries were due to maternal request, planned, or scheduled (Glance, Dick, Osler, & Mukamel, 2006; Goldman, Chu, Bacchetti, Kruger, & Bindman, 2015). These differentials were known to bias risk adjusted estimates of baseline comparisons or hospital ranks. For provider comparisons, or studies of trends over time, some indicators suffered from small sample size due to low volume admissions (Bardach, Chien, & Dudley, 2010) and more robust statistical techniques were required (Moineddin, Matheson, & Glazier, 2007).

Robson 10 Group Classification System

Robson 10 Group Classification System was introduced around 1980 by Michael Robson of the United Kingdom. Its system was used at the hospital level to audit obstetrical management prospectively and to improve the quality of cesarean and vaginal delivery procedures (Robson et al., 1996). Since then, over 75 studies have applied its system and varied according to research design, unit of analysis, and geographical place (Betran, Vindevoghel, Souza, Gulmesoglu, & Torloni, 2014). A peer review of Robson's system in comparison to 26 other cesarean delivery risk assessment methods identified it as ranking the highest according to attributes of use, reproducibility, and ability to follow prospectively and classify patients before the cesarean procedure (Torloni et al., 2011). Noted weaknesses involved its inability to identify reasons for the cesarean delivery according to clinical or nonclinical indications (Torloni et al., 2011).

Robson Groups 1, 2, and 5 were similar to cesarean indicators already proposed for U.S. population-based monitoring or hospital quality improvement programs (NCHS, 2012). Often excluded from the U.S. system was the systematic monitoring of other Robson groups specific to abnormal lies (Groups 6, 7, and 9), multiple pregnancy (Group 8), preterm cases (Group 10), and multiparous events (Groups 3 and 4). A proportion of these latter groups may have represented inappropriate cesarean delivery utilization and contributed to rising health care costs or poor pregnancy outcomes (Kabir et al., 2004; Queenan, 2011; Scott, 2011; Srinivas et al., 2010). U.S. cesarean delivery indicators often excluded measures for labor induction and metrics for cesarean delivery events occurring before spontaneous delivery. Both practices were identified as contributing to inappropriate cesarean delivery utilization, especially when there were no clinical indications or for deliveries of infants less than 39 weeks gestational age (Donovan et al., 2010; Ehrenthal, Hoffman, Jiang, & Ostrum, 2011; Zhang et al., 2010).

Robson's Classification System used stratification methods for the examination of population-based (Betran et al., 2007; Denk et al., 2006; Zhang et al., 2010), hospital level, or group specific differences in rates. Most studies were descriptive and generated unadjusted rates for comparison within a hospital (Costa, Cecatti, Souza, Milanez, & Gulmezogulu, 2010; Florica, Stephansson, & Nordstrom, 2006; McCarthy, Rigg, Cady, & Cullinane, 2007; Scarella, Chamy, Sepulveda, & Belizan, 2010) between hospitals

(Brennan et al., 2009; Rasmussen, Pedersen, Wilken-Jensen, & Vejerslev, 2000), or according to patient risk (Howell, Johnston, & Macleod, 2009). A limited number of studies applied risk adjustment methods as a means to control for confounding (Betran et al., 2007), or used advance statistical techniques to model associations (Allen, Baskett, & O'Connell, 2010; Fisher, LaCoursierre, Barnard, Bloebaum, & Varner, 2005; Maso et al., 2013).

The majority of Robson studies used the cephalic concept as a measure of presentation at birth, with only one identified study using the vertex measure (Stavrou, Ford, Shand, Morris, & Roberts, 2011) and in alignment with U.S. population-based cesarean delivery methods (Coonrod et al., 2008; Kahn et al., 2009; Main et al., 2006). A few studies adapted Robson's tool and narrowly examined a few groups (Fischer et al., 2005), aggregated groups based on differing research purposes (Denk et al., 2006; Zhang et al., 2010), or in relation to maternal or infant health outcomes (Homer, Kurinczuk, Spark, Brocklehurst, & Knight, 2007). All approaches demonstrated the tool's flexibility for research, surveillance, and quality improvement.

Among U.S. population-based studies having applied Robson's Classification, live birth files were used to examine retrospectively group specific cesarean delivery rates in the reporting of state level aggregated rates (Denk et al., 2006). At least two hospital studies existed with case ascertainment via the hospital medical chart (Fischer et al., 2005) or electronic health record (Zhang et al., 2010). These latter studies demonstrated that more specific information was ascertained for classifying patients according to clinically relevant indications. Population-based studies (Denk et al., 2006;

Zhang et al., 2010) precluded the examination of hospital specific rates and their comparison, or determined why cesarean deliveries varied within or between hospitals. No U.S. population-based study, having examined Robson Classification indicators, employed population-based, data linking strategies and the use of administrative data sources for case ascertainment.

Provider Profiling

U.S. provider profiling emerged in 1987 with the Health Finance Administration's release of hospital level mortality rates of coronary artery bypass grafts (Delong et al., 1997; Normand, Glickman, & Gatsonis, 1997). Since then, U.S. federal, state, and local initiatives demonstrated healthcare profiling initiatives. At the federal level, the U.S. Department of Health and Human Services' (DHHS) Medicare Hospital Compare Report was published through the internet and disseminated hospital based indicators for comparison to subnational and national benchmarks (DHHS, 2012). Whereas, the state of Illinois, through its Hospital Report Card Act (Public Act 93-0563, 2004), mandated the public disclosure of an Illinois Hospital Report Card and Consumer Guide to Health Care, which was accessible through its internet site (Illinois Department of Public of Health, 2012). California, Colorado, Massachusetts, New York, and Utah were national leaders in advancing provider report cards for the examination of hospital performance (Racz & Sedransk, 2010; Shahian et al., 2005; State of California, 2012).

Provider profiling was an emerging health services science, and its practice faced gradients of acceptance and proven effectiveness. Cardiovascular health was an early adopter (Delong et al., 1997; Normand, Glickman, & Gatsonis, 1997), and over time, methods were advanced through consensus, findings from research, and translation into evidence based practice (Krumholz et al., 2005; Shahian et al., 2011). Like the cesarean delivery procedure, cardiovascular health was motivated by high cost of care and the high volume of medicine.

Profiling techniques improved the accuracy of report cards by employing robust methods for the measurement of structure, process, and outcomes associated with health services (Donebedian, 2002). Approaches used statistical techniques to estimate standards and observed-to-expected rates for the comparison of providers to benchmarks and review of outliers that deviated from expectation (Ash et al., 2012). Risk adjustment methods modelled variation in health outcomes across providers based on differences in baseline covariates (Iezzoni, 1997). In the examination of outliers, statistical methods were employed to refine the understanding of practice patterns that reflected underuse, overuse, or misuse of service utilization. Risk adjustment analytic techniques that were conceptually relevant and balanced by parsimonious approaches for routine use remained a challenge.

The hospital was often used as the unit of analysis, with other studies examining the performance of physicians or nursing homes. Typically, hospital profiling methods compared all institutions with each other (Fantini et al., 2006; Paranjothy, Frost, & Thomas, 2005) for the comparison to an average benchmark score. Measurable differences in health outcomes were often attributed to variation in provider quality, patient severity of illness, accuracy of profiling methods, random error, or statistical error (Fung et al., 2010).

An early U.S. maternity study introduced provider profiling through a statewide, cross-sectional study of 1994 California hospital discharge summary data and the examination of interhospital effects of cesarean and vaginal deliveries on the length of stay outcome (Leung, Elashoff, Rees, Hasan, & Legorreta, 1998). These emerging health services methods benefitted from novel multilevel statistical techniques earlier introduced by the fields of psychology and education (Raudenbush & Byrk, 1992). Compared to more conventional statistical techniques, multilevel methods proved statistically robust in modeling health services data, typically characteristic of nested features, for an understanding of intra- and inter-provider effects on outcomes. Multilevel statistical techniques were also enhanced by seminal work in the area risk adjustment for health services outcomes (Iezzoni, 1997). Without random assignment, risk adjustment methods were employed to control for confounding of health outcomes and the reduction of biased effects for the observational design.

Hospital Profiling of the Cesarean Delivery Outcome

At least eight U.S. studies demonstrated hospital profiling of the cesarean delivery procedure. Study designs varied according to time period, sampling frame, statistical technique, and cesarean delivery indicator type. The majority of studies were retrospective in design, having examined the total cesarean delivery rate, and at least two studies presented hospital profiles for the nulliparous-term-vaginal-singleton (NTVS-CS) rate or a metric for low risk cesarean section patients (Coonrod et al., 2008; Main et al., 2006). Live birth files (Bailit et al., 1999; Bailit & Garrett, 2003; Coonrod et al., 2008), hospital discharge summary data (Leung et al., 1998), or linked hospital live birth files

(DiGiuseppe et al., 2001; Keeler et al., 1997), were used for case ascertainment. Most designs included the study of regional (Aron et al., 1998; DiGiuseppe et al., 2001; Kritchevsky et al., 1999; Main et al., 2006) or statewide examinations (Bailit et al., 1999; Coonrod et al., 2008; Keeler et al., 1997; Leung et al., 1998). Typical statistical techniques used in risk adjustment of the cesarean delivery outcome and in the comparison of providers include logistic regression (Aron et al., 1998; Bailit & Garrett, 2003; Coonrod et al., 2008; Glantz, 1999; Kritchevsky et al., 1999) or indirect standardization (Glantz, 1999). At least one study applied multilevel statistical techniques for modeling patients nested within hospitals and in the study of cesarean delivery variation (Leung et al., 1998).

The majority of U.S. hospital profiling studies of the cesarean delivery outcome were dated and using informational sources from up to twenty years ago. Although multilevel statistical methods were endorsed by other research areas for profiling methods, a limited number of U.S. cesarean delivery studies applied these more advanced statistical techniques. Only two studies examined hospital profiles using the more refined NTVS-CS indicator (Coonrod et al., 2008; Main et al., 2006). Even then, risk adjustment methods ranged from the more parsimonious, including only age (Main et al., 2006), the comparative modeling of mother’s age and infant birth weight, or clinical indications for the cesarean delivery outcome (Coonrod et al., 2008).

Critical to hospital profiling methods involved the accurate development of risk adjustment methods. For the cesarean delivery outcome, a standard risk adjustment method did not exist. Rather, logistic regression was typically used to model a complex set of variables, measuring patient sociodemographics, preexisting medical conditions, medical risk factors of the index pregnancy, or severity of illness. Among cesarean delivery risk adjustment studies reviewed, limited attention was given to comparing differing risk adjustment methods according to covariate selection or statistical technique for an understanding of selection bias.

For regional or statewide studies using live birth data (Coonrod et al., 2008), cesarean delivery risk adjustment methods were limited to the available informational source, and linkages with other population-based data sources could have improved upon covariate selection and reduced selection bias (Kahn et al., 2009; Lydon-Rochelle, Holt, Cardenas, et al., 2005; Lydon-Rochelle, Holt, Nelson, et al., 2005; Stivanello et al., 2011). For example, U.S. population-based linked files ascertained more refined cesarean delivery classifications of disease when compared to the live birth file, alone (Kahn et al., 2009). Yet, absent from administrate data sources were more accurate covariates or refined specifications for maternal request (Barber et al, 2011), Bishop score (Zhang et al., 2010), or a present on admission flag to distinguish baseline covariations from those that were hospital acquired or complications due to hospital procedures (Glance, et al., 2006).

Without risk adjustment, potential confounding was ignored, and it was assumed that there was no difference in provider treatment (Iezzoni, 1997). However, providers varied according to patient case mix, organizational or background characteristics of physicians, and the health systems in which they practiced (Burns et. al, 1995). Selection bias also threatened risk adjustment methods due to the quality of variable used, when known confounders were excluded, or when unmeasured confounders were excluded or unknown (Huang, Dominici, Frangakis, et al., 2005). Studies demonstrated provider profiles having applied risk adjusted rates produced improved estimates in health outcomes when compared to unadjusted rates (Mukamel et al., 2008; Normand et al., 1997). On the other hand, a more recent study showed marginal differences between crude and risk adjustment rates models for the NTVS-CS procedure (Stivanello et al., 2012). Further examination of risk adjustment comparative methods assisted in determining whether findings were attributed to covariate selection, statistical technique, or both (Huang, Dominici, Frangakis, et al., 2005).

To advance provider profiling of the cesarean delivery outcome, multilevel statistical technique and risk adjustment of outcomes were proposed (Krumholz et al., 2005). Studies that compared multilevel, risk adjustment models with the same covariates modeled with logistic regression demonstrated varying effects on provider ranks, statistical outliers, point estimates, or coefficients (Alexandrescu, Jen, Bottle, Jamare, Aylin, 2011; Austin, Tu, Alter, 2003; D’Errigo, Tosti, Fusco, Perucci, & Seccareccia, 2007; Glance et al., 2006; Huang, Dominici, Frangakis, et al., 2005; Shahian et al., 2005). Multilevel statistical methods were known to produce more conservative point estimates and confidence intervals when compared to logistic regression (Huang, Dominici, Frangakis, et al., 2005). This was attributed to the multilevel statistical capability of partitioning variation of the nested design; the pooling of data across all providers and shrinkage toward a grand mean; and the mechanics of fixed and random effects, with the latter modeling covariance across all provider, whereas logistic regression was limited to fixed effects, only (Raudenbush & Byrk, 2002).

Propensity Score Methods

To advance profiling methodologies, propensity methods were proposed. Methods were applied to the observational design to balance baseline covariates for the treatment and control condition and suggested an alternative to the randomized control trial (West et al., 2008). First introduced in 1983 (Rosenbaum & Rubin, 1983), propensity score methodologies were advanced through epidemiology, health services, and economic research as a valid tool for balancing covariates for two group (Rosenbaum & Rubin, 1983) and multinomial comparisons (Huang, Frangakis, Dominici, Diette, & Wu, 2005; Imbens, 2000; Shahian & Normand, 2008; Spreeuwenberg et al., 2010). Emerging studies that compared health outcomes based on propensity score models versus randomized controlled trials demonstrated comparable effects, suggesting reduced bias and model precision as demonstrated by the observational design (Kuss, Legler, & Borgermann, 2011; Smeeth, Douglas, Hall, Hubbard, & Evans, 2008; Steiner, Shadish, Cook, & Clark, 2010).

The literature highlighted four propensity score conditioning methods and included stratification, regression, inverse probability of treatment weighting (IPTW), and matching, with varying performance in model effects when applied (Austin, 2011). Observational studies comparing differing propensity score methods and holding variables constant demonstrated differences in risk adjusted outcomes (Austin & Mamdani, 2006) or provider rankings (Huang, Frangakis, Dominici, et al., 2005; Shahian

& Normand, 2008). Findings suggested the type of covariate selection affected bias more than the choice of the specific propensity score statistical technique (Huang, Frangakis, Dominici, et al., 2005; Steiner et al., 2010). On the other hand, Monte Carlo simulation studies demonstrated that model performance varied by propensity score type and variable specification and unmeasured confounding of important variables resulted in a significant imbalance between the treatment and control groups (Austin, Grootendorst, Anderson, & Norman, 2007). Simulation findings also demonstrated matching and inverse weighting produced unbiased estimates as compared to stratification and propensity score risk adjustment (Austin, 2011). Propensity score estimates were also affected by observed sample size (Belitser et al., 2011) and the number of variables entered into models. Findings suggested biased estimates for events with less than 10 events (Rassen, Glynn, Brookhart, & Schneeweiss, 2011) and more robust estimates for models having large sample sizes (Belitser et al., 2011).

In a comparative study, multivariable logistic regression, propensity score matching, propensity score adjustment, and propensity score weighting were applied to an observational design for the measure of tissue plasminogen activator on death of 6,269 ischemic stroke patients listed in a German stroke registry (Kurth, Walker, Glynn, Chan, & Gaziano, 2005). Findings suggested the crude risk adjustment model (*OR* = 3.35, 95% CI [2.28, 4.91]) differed from the multivariable model (*OR* = 1.93, 95% CI [1.22, 3.06]) in both point estimates and confidence intervals. When matched on the propensity score (*OR* = 1.17, 95% CI [0.68, 2.00]), this model also differed when compared to logistic regression adjusted with propensity score (*OR* = 1.53, 95% CI [0.95, 2.48]).

A comparative evaluation of differing propensity score methods, using Monte Carlo simulation for the estimated effect on binary outcomes, demonstrated model variation (Austin, 2008) as well. For the covariate adjustment model using the estimated propensity score, findings suggested the treatment effects were biased towards the null, irrespective of whether the true propensity score or other confounders were included in statistical models (Austin, 2008). For models using propensity score conditioning on matching, findings suggested model bias was less affected when the true odds ratio was less than or equal to one (Austin, 2008).

More recently, propensity score risk adjustment methods were applied for the examination of elective repeat cesarean delivery versus spontaneous trial of delivery after prior cesarean delivery (Gilbert et al., 2012). This was a retrospective study design and used 1999-2002 registry data from nineteen U.S. clinical centers. Propensity score matching was used to balance baseline patient demographic and clinical conditions to reduce confounding. Conditional logistic regression methods were applied to model the cesarean delivery association with the generation of odd ratios and confidence intervals. Notably absent were comparative statistical methods to understand whether the study design employed appropriate selection of propensity score covariates, balance was achieved between the exposure and control groups, or if statistical methods produced unbiased effects (Austin, 2008).

To advance provider profiling, propensity score methods were proposed for risk adjustment models (Shahian & Normand-Lise, 2008) and in the generation of rank (Huang, Frangakis, Dominici, et al., 2005). Propensity scores had been used to balance covariates across multiple providers, acknowledging patient selection into a particular provider (or treatment) was based on nonrandom assignment (Rosenbaum & Rubin, 1983). Achieving covariate balance across all provider units aimed to reduce biased estimates in health outcomes due to confounding. The ability to balance covariates across providers required careful statistical technique or outcomes faced bias due to methods associated with propensity score model selection, covariate specification, or sample size. Models assumed all known confounders were measured and with no unmeasured confounding associated with the treatment and outcome (Rosenbaum & Rubin, 1983).

Huang, Frangakis, Dominici, et al., 2005 (2005) applied Imbens (2000) propensity score methodology for multinomial comparisons across twenty physician groups in the study of asthma treatment in children. The analytic framework compared the performance of physician rankings derived from three differing risk adjustment methods, including multilevel statistical technique, propensity score risk adjustment, and having no risk adjustment. Covariates used for matching included patient endogenous variables, including age, race, education, severity of illness, those present at baseline, and not easily modifiable given treatment. Stratification methods were used to group patients into five separate quintiles based on comparable propensity scores for the balanced design. In the comparison of physician rankings, propensity score risk adjustment demonstrated shifts in rank order when compared to multilevel statistical methodologies.

Notably absent from the multilevel modeling were physician level variables. Huang, Dominici, Frangakis, et al. (2005) posited the latter were exogenous variables, and should be excluded, because these effects explained away variation which the study design aims to measure differences in quality.

Cardiovascular health built upon this work and applied risk adjusted propensity scores for the examination of performance ranking of 14 Massachusetts hospitals and the utilization of coronary artery bypass surgical (CABG) procedure (Shahian & Normand, 2008). Multinomial regression modeling was used to estimate propensity scores using selected CABG risk factors. For each patient, 14 differing probabilities were estimated, reflecting the distribution of hospitals sampled and summing to a total score of one. Logistic regression modeling was then used to estimate CABG mortality rate. Findings demonstrated significant differences among hospitals according to demographic, comorbid, or severity of illness factors.

Summary

The literature review demonstrated the majority of hospital profiling studies for the cesarean delivery outcome were dated. Since these early studies, prevailing provider profiling methodologies endorsed the use of multilevel statistical techniques to model the structured characteristics of patients nested within hospitals in the study of outcomes. Likewise, there were a limited number of population-based cesarean delivery studies that used risk adjustment and statistical technique comparative methods. No study was identified as having examined hospital profiles for the cesarean delivery procedure with applied propensity score methods. No study was found having applied propensity score matching with stratification for conditioning on the cesarean delivery outcome. No U.S. study was identified having extracted Robson 10 Group indicators from population-based linked datasets. Although a number of organizations and experts continued called for the reduction of inappropriate utilization, there were limited population-based quality improvement approaches for ongoing, systematic monitoring and timely response. The next chapter describes proposed methodologies to advance hospital profiling and propensity score methodologies for the cesarean delivery procedure.

Chapter 3: Research Method

Introduction

The purpose of this research was to create a statewide hospital profiling and propensity score matching methodology for the study of variations in the cesarean delivery procedure. The research methodology applied data linking, hospital profiling, and propensity score matching in the examination of the cesarean delivery outcomes.

Research Design and Rational

This was a population-based, retrospective observational design using 2012 linked data for the state of Georgia. The cesarean delivery procedure was defined as the dependent variable and Robson 10 Group Classification indicators, clinical conditions, sociodemographic characteristics, and patient level propensity scores were selected as independent variables. Logistic regression and hierarchical generalized linear modeling were used to examine variation in hospital ranks, model effects, and strength of association. Propensity score matching with stratification was applied to the observational design for the study of the association of model effects on the cesarean delivery procedure.

Methodology

Study Population

The study population included maternal live births from the state of Georgia as represented by a 2012 linked dataset. The 2012 linked dataset included extracted records from the 2012 Georgia live birth file and 2012 Georgia hospital discharge summary file based on standard case definitions and code sets. The Georgia hospital discharge summary data was produced by the Georgia Hospital Association and included all nonfederal hospital admissions for the state of Georgia. This data file recorded patient sociodemographic, comorbidity, complications, severity of illness scores, and outcomes of hospital admission. Eligible cases included vaginal and cesarean deliveries listed for the period, January 1, 2012 to December 31, 2012. Fetal death, antepartum, and postpartum events unrelated to delivery, abortions, ectopic pregnancy, molar pregnancy, and other abnormal products of conception were excluded from the hospital discharge summary file prior to data linking. The 2012 Georgia live birth file was produced by the Georgia Department of Public Health. This file listed sociodemographic, comorbidities, complications, and outcomes of pregnancy for each live birth record. Eligible cases included Georgia live births for the period January 1, 2012 to December 31, 2012. Excluded cases were ones classified as home births, delivered at a federal hospital, and out of state births.

Dependent Variable

The dependent variable was the cesarean delivery outcome. Table 2 summarizes the standard codes sets used to extract the dependent variable from the 2012 linked file. The dependent variable was a binary outcome measured by cesarean delivery = 1 and vaginal delivery = 0. The cesarean delivery and vaginal variables were extracted from the 2012 linked dataset using the hospital discharge summary's Diagnosis Related Group (DRG) codes, International Classification of Disease, Ninth Revision, clinical modification (ICD-9-CM) procedure codes, and ICD-9-CM diagnosis codes. Live birth file standard codes sets for the vaginal and cesarean method of delivery were used as external validation.

Table 2

Hospital Discharge Summary File Extraction Code Sets for Method of Delivery

Method of delivery	DRG code	ICD-9-CM procedure code	ICD-9-CM diagnosis code
Singleton vaginal	767, 768, 774, 775	72.0 – 72.99	650.00, V27.0
Multiple vaginal	767, 768, 774, 775	72.0 – 72.99	V27.2, V27.5
Multiple vaginal mixed	767, 768, 774, 775	72.0 – 72.99	V27.3, V27.6
Singleton cesarean	765 - 766	74.0 – 74.99	650.00, V27.0
Multiple cesarean	765 - 766	74.0 – 74.99	V27.2, V27.5
Multiple cesarean mixed	765 - 766	74.0 – 74.99	V27.3, V27.6

Independent Variables

Independent variables included 10 Robson Group indicators, 16 clinical conditions, two sociodemographic characteristics, and two patient level propensity scores. Variable selection was guided by cesarean delivery risk adjustment and hospital profiling literature.

Table 3

Robson 10 Group Classification Extraction Rules by File Type

	2012 live birth file	2012 hospital discharge summary file		
Robson concept	Data code	MS-DRG code	ICD-9-CM procedure code	ICD-9-CM diagnosis code
Cesarean	MOD = 2	765 - 766	740, 741, 742, 744, 749, 7499	669.70 - 669.71
Vaginal	MOD =1	767, 768, 774, 775	720, 721, 7221, 7229, 7231, 7239, 724, 726, 7251-7254, 7271, 7279, 7322, 7359	
Cephalic				650.00
Abnormal lie				652.3 - 652.93
Breech			725 - 7254	652.1 - 652.23, 669.60 - 669.61
Plurality	Singleton = 1, Other = 0			
Uterine scar				654.2-654.23
Parity	Nulliparous = 1, Multiparous = 2			
Preterm	Preterm = 1, Other = 0			
Induction	Induction =1, Other= 0		7301, 731	659.00 - 659.11, 660.60 -660.61, 661 - 661.23, 662.00 - 662.21

Robson 10 Group Classification was defined by mutually exclusive indicators, ranging from Group 1 to Group 10. Robson Groups were extracted from the 2012 linked file using variables from the live birth file, the hospital discharge summary record, or both. Table 3 and Table 4 summarize the key variables required to extract the 10 Robson Groups, by standard code sets, and whether the live birth file or hospital discharge file were required.

Clinical conditions were selected according to their association with the cesarean delivery outcome, application to previous cesarean risk adjustment, hospital profiling methods, and data quality of the resulting 2012 linked file. Table 5 includes 16 variables selected as clinical characteristics, whether they were extracted from the hospital discharge summary record or the live birth record, and the standard code sets used. Cesarean delivery clinical characteristics varied according to severity with diabetes, eclampsia, RH sensitization, HIV, dystocia, fetal distress, and fetal abnormal heart considered severe conditions (Bailit & Garrett, 2003). Moderate medical conditions included polyhydramnios, oligohydramnios, dystocia, hypertension, and fetal abnormality (Bailit & Garrett, 2003). Mild medical conditions included anemia and genital herpes (Bailit & Garrett, 2003).

Two sociodemographic variables were selected and included age and insurance. Age was extracted from the live birth file, and insurance was ascertained from the hospital discharge summary file. Race was excluded because this is an endogenous characteristic and its direct effect on change in quality may not be readily manifested (Huang, Dominici, Frangakis, et al., 2005). Also, the race variable listed in the hospital discharge summary file and live birth file represented high incongruence due to the response categories listed.

Two patient level propensity scores were generated from logistic regression models. The method of delivery was defined as the outcome variable and 10 Robson Groups, 16 clinical conditions, and two SES variables were modelled as risk adjustors. The resulting model produced a predicated probability for each patient and was defined as the true propensity score. Logarithmic (LN) transformation was applied to each patient’s true propensity score and resulted in the LN propensity score.

Table 4

Standard Code Sets Used in Creating Robson 10 Group Composite Indicators

Robson Group	Description	Standard code set
Group 1	Nulliparous, single cephalic, ≥ 37 weeks	parity = 1; plurality = 1; cephalic = 1; preterm = 0; induction = 0; breech = 0; abnormal lie = 0; uterine scar = 0
Group 2	Nulliparous, single cephalic, ≥ 37 weeks, induced	parity = 1; plurality = 1; cephalic = 1; preterm = 0; induction = 1; breech = 0; abnormal lie = 0; uterine scar = 0
Group 3	Multiparous (no previous CS), single cephalic, 37 ≥ weeks, in spontaneous labor	parity = 2; plurality = 1; cephalic = 1; preterm = 0; induction = 0; breech = 0; abnormal lie = 0; uterine scar = 0
Group 4	Multiparous (no previous CS), single cephalic, 37 ≥ weeks, induced	parity = 2; plurality = 1; cephalic = 1; preterm = 0; induction = 1; breech = 0; abnormal lie = 0; uterine scar = 0
Group 5	Previous uterine scar, single cephalic, 37 ≥ weeks	uterine scar =1; plurality = 1; cephalic = 1; preterm = 0; breech = 0; abnormal lie = 0
Group 6	All nulliparous breeches	parity = 1; breech = 1; abnormal lie = 0; cephalic = 0
Group 7	All multiparous breeches, with or without previous uterine scar	parity = 2; breech = 1; abnormal lie = 0; cephalic = 0; uterine scar = 1; uterine scar = 0
Group 8	All multiple pregnancies, with or without previous cesarean delivery	plurality = 0; uterine scar = 1; uterine scar = 0
Group 9	All single pregnancies with abnormal lies, with or without uterine scar	plurality = 1; abnormal lie = 1; breech = 0; cephalic = 0; uterine scar = 1; uterine scar = 0
Group 10	All single cephalic, ≤ 36 weeks, with or without previous cesarean delivery	plurality = 1; cephalic = 1; preterm = 1; uterine scar = 1; uterine scar = 0; breech = 0; abnormal lie = 0

Table 5

ICD-9-CM Standard Codes for Extracting Maternal Clinical Conditions from the Hospital Discharge Summary File

Clinical condition	ICD – 9 – CM diagnosis code
Diabetes	248.8, 250, 648.0,
Hypertension	401.0 – 405.9, 642.0 – 642.59, 642.7 – 642.79, 642.9 – 642.93
Eclampsia / Pre-eclampsia	642.40, 642.1, 643.43, 642.5, 642.60 - 642.69, 642.71, 642.73
HIV	042, V08
Oligohydramnios	658. 01 – 657.03
Polyhydramnios	657.00 – 657.03
Fetal abnormality	655.01 – 655.91
Antepartum	641.00 – 641.03, 641.10 – 641.13, 641.20 – 641.23, 641.30 – 641.33, 641.80 – 641.83, 641.90 – 641.93, 668.80
Fetal distress	656.2 – 656.33, 659.7 – 650.73
Anemia	280.0 – 282.3, 282.8 – 285.0, 285.2 – 285.9, 648.20 – 648.24
Dystocia	653, 660.01 – 660.91, 661.01 – 661.91, 662.01 – 662.31
Fetal anomaly	655.01 – 655.91
RH	656.1 – 656.13
PROM	658.11 – 658.31
Genital herpes	54.1 – 54.19, 282.4 – 282.79
Fetal heart	659.70 – 659.73

Deterministic Data Linking

SPSS Version 21.0 was used to conduct deterministic data linking of the 2012 Georgia hospital discharge summary file and the 2012 Georgia live birth file. Key variables used in data linking algorithms included a 15-digit unique ID, facility code, mother's date of birth, live birth event date, hospital discharge date, and hospital admission date.

Table 6

Deterministic Matching Variable by Data Linking Algorithm

Variable	Algorithm 1	Algorithm 2	Algorithm 3
15-digit unique ID*	x		
Concatenated 15-digit unique ID*		x	
Concatenated 14-Digit unique ID*			x
Facility code*		x	x
Mother's date of birth*		x	x
Event date #		x	x
Hospital admission date+		x	x
Hospital discharge date+		x	x

Note. * = matching variable from Hospital Discharge Summary File and Live Birth File, # = validation variable from the Live Birth File, + = validation variable from the Hospital Discharge Summary File.

Table 6 describes the data linking algorithms applied for the production of the 2012 Georgia linked file. Algorithm 1 used deterministic data linking methods to match patient level data from the 2012 Georgia hospital summary data file and the 2012 Georgia live birth file using a 15-digit unique ID. Exact matches were stored as a linked dataset and unlinked files were reused in the second data linking algorithm. Data linking Algorithm 2 used a different set of variables including a concatenated 15-digit unique ID, facility code, and mother's date of birth. In addition, the event date from the live birth file, or infant date of birth, was used to verify against the hospital discharge summary hospital admission and discharge date. Event dates falling on or near the admission date were considered as an exact match. Matched cases from the second data linking algorithm were appended to the linked dataset and unlinked files were reused in Algorithm 3. Lastly, Algorithm 3 matched on a concatenated 14-digit unique ID, facility code and mother's date of birth. Similar to Algorithm 2, the event date was verified using the hospital admission and discharge dates. Matched records were appended to the linked file and unlinked records were discarded.

Data Analysis Plan for Individual Research Questions

The data analysis plan detailed the research methodology for seven research questions, specific to deterministic data linking, Robson 10 Group Classification extraction methods, hospital profiling, and propensity score matching.

Robson 10 Group Classification Rates

Research Question 1 evaluated the accuracy of extracting Robson 10 Group Classification indicators (Robson et. al., 1996) from the 2012 Georgia linked data file. No U.S. published study was identified that used population-based linked datasets for the extraction and that used Robson 10 Group indicators as a risk adjustor in multivariate analyses. The expectation was that the Robson 10 Group rates ascertained from the 2012 Georgia linked file would differ in their mean comparisons.

1. RQ 1: Is there a difference in the mean cesarean delivery rate among the Robson 10 Groups?

$H_0$1: There is no difference in the mean cesarean delivery rate among the Robson 10 Groups.

H_a1: There is a difference in the mean cesarean delivery rate among the Robson 10 Groups.

For each of the 10 Robson indicators, three descriptive statistics were produced (Robson et al., 1996). The relative size was defined as the total number of group specific deliveries and included both vaginal and cesarean events. The cesarean delivery rate was estimated for each Robson Group and defined as the group specific number of cesarean deliveries divided by the total number of vaginal and cesarean deliveries. The absolute contribution estimated the group specific number of cesarean deliveries compared to the total cesarean delivery rate. SPSS 21 was used to generate one-way ANOVA and Tukey multiple comparison tests.

Hospital Profiling

Research Questions 2 to 5 modelled risk adjustment and statistical techniques on the cesarean delivery outcome for the generation of model effects and hospital rankings (Table 6). Comparative risk adjustment and statistical methods were proposed, given their limited application to hospital profiling of the cesarean delivery outcome to date. The expectation was that cesarean delivery model effects and hospital ranks would vary by risk adjustment model and statistical technique.

2. RQ 2: Is there a difference in cesarean delivery risk adjustment effects when

comparing logistic regression versus hierarchical generalized linear models?

$H_0$2: There is no difference in cesarean delivery risk adjustment effects when comparing logistic regression versus hierarchical generalized linear models.

H_a2: There is a difference in cesarean delivery risk adjustment effects when comparing logistic regression versus hierarchical generalized linear models.

3. RQ 3: Is there a difference in cesarean delivery risk adjustment effects when comparing logistic regression propensity score versus hierarchical generalized linear propensity score models?

$H_0$3: There is no difference in cesarean delivery risk adjustment effects when comparing logistic regression propensity score versus hierarchical generalized linear propensity score models.

H_a3: There is a difference cesarean delivery risk adjustment effects when comparing logistic regression propensity score versus hierarchical generalized linear propensity score models.

4. RQ 4: Is there a difference in cesarean delivery hospital ranks when comparing logistic regression versus hierarchical generalized linear models?

$H_0$4: There is no difference in cesarean delivery hospital ranks when comparing logistic regression versus hierarchical generalized linear models.

H_a4: There is a difference in cesarean delivery hospital ranks when comparing logistic regression versus hierarchical generalized linear models.

5. RQ 5: Is there a difference in cesarean delivery hospital ranks when comparing logistic regression propensity score risk adjustment versus hierarchical generalized linear propensity score risk adjustment models?

$H_0$5: There is no difference in cesarean delivery hospital ranks when comparing logistic regression propensity score risk adjustment versus hierarchical generalized linear propensity score risk adjustment models.

H_a5: There is a difference in cesarean delivery hospital ranks when comparing logistic regression propensity score risk adjustment versus hierarchical generalized linear propensity score risk adjustment models.

Hospital profiling approaches aimed to test statistical techniques and risk adjustment variable selections in order to understand their effect on hospital ranking of the cesarean delivery outcome and measures of association. Logistic regression and HGLM statistical techniques were known to differ in their approach and in modeling fixed and random effects. Logistic regression models were selected because these were typical of reproductive health studies that examined variation in the cesarean delivery outcome. A 2-level HGLM model was selected because these models included specifications for analyzing studies with a binary outcome and are known to account for the nesting of patients within hospitals and modelled fixed and random effects (Raudenbush & Byrk, 1996).

Table 7

Hospital Profiling of the Cesarean Delivery Outcome by Model Type

Statistical technique	Risk adjustment model
Logistic Regression Model 1 (LR Null)	Dummy variables for 85 hospitals
Logistic Regression Model 2 (LR Full)	Robson 10 Groups + 16 clinical conditions + 2 SES + dummy variables for 85 hospitals
Logistic Regression Model 4 (LR PS)	Robson 10 Groups + 16 clinical conditions + 2 SES + propensity score + dummy variables for 85 hospitals
Logistic Regression Model 4 (LR LN)	Robson 10 Groups + 16 clinical conditions + 2 SES + LN propensity score + dummy variables for 85 hospitals
HGLM Model 5 (HGLM Null)	Level-2: Dummy variables for 85 hospitals
HGLM Model 6 (HGLM Full)	Level-1: Robson 10 Groups + 16 clinical conditions + 2 SES
	Level-2: dummy variables for 85 hospitals
HGLM Model 7 (HGLM PS)	Level-1: Robson 10 Groups + 16 clinical conditions + 2 SES + propensity score
	Level-2: dummy variables for 85 hospitals
HGLM Model 8 (HGLM LN)	Level-1: Robson 10 Groups + 16 clinical conditions + 2 SES characteristics + LN propensity score
	Level-2: dummy variables for 85 hospitals

SPSS Version 21.0 was used to create four logistic regression models for the comparative evaluation of model effects and hospital ranks (Table 7) (Appendix A). Model 1, or the null model, was considered the crude approach and did not include any risk adjustors, only hospital dummy variables. Models 2 through 4 included Robson Groups, clinical, SES, and propensity score risk adjustors, plus dummy variables for 85 hospitals. Ten dummy variables were created for the Robson Groups with Group 5 defined as the referent group because of an anticipated observed rate greater than expected. Dummy variables were created for each of the 16 clinical characteristics. Dummy variables were created for three insurance variables, and public insurance was identified as the referent group. Five dummy variables were created for maternal age and women 25 to 30 years of age were defined as the referent group. Eighty-five hospital dummy variables were created with Hospital 1 defined as the referent facility. The referent hospital represented a medium to high volume delivery hospital with one of the lowest cesarean delivery rates in the state of Georgia.

Hosmer-Lemeshow tests guided the evaluation of logistic regression model fit (Hosmer & Lemeshow, 1989). Model effects were reported using the odds ratio (*OR*), 95% confidence intervals (CI), and p values. The odds ratio was a standard point estimate used in reproductive health hospital profiling and risk adjustment studies, assisting in the evaluating the strength of association of measured effects in relation to the study outcome (Handler, 1998). Confidence intervals were reviewed according to upper and lower limits and in relation to the odds ratio. Statistical significance was set at an alpha level less than 0.05.

Four hierarchical generalized linear models (HGLM) were developed and followed a similar rational to the logistic regression methodologies (Table 7) (Appendix B). HLM Version 7 software was used to model a 2-level HGLM binomial model for the cesarean delivery outcome. Patient level risk adjustors were added to Level-1 and hospital dummy variables to Level-2. The method of estimation used a full maximum likelihood with Laplace iterations. Level-1 variables were centered on the group mean, and Level-2 covariates were grand mean centered (Raudenbush, Byrk, Cheong, Congdon, & Toit, 2004). For each HGLM model, the unit specific model with the logit link function and the population average model were reviewed. Unit specific random effects for each model were reviewed for an understanding of variance components and tests of significance.

Standardized model effects, odds ratios, and hospital ranks generated from the logistic regression and HGLM models were reviewed according to measures of association and statistical tests (Alexandrescu, Jen, Bottle, Jamar, Aylin, 2011; Austin, Tu, Alter, 2003; Kurth, 2005). Maternal child health analysis guidelines were also used to describe the magnitude of association (Handler, 1998). A point estimate with a strong association was defined as an odds ratio ranging from 3.00 - 10.00 or 0.01 - 0.33. A moderate association included an odds ratio ranging from 0.34 - 0.67 or 1.50 - 2.99. A weak association was defined as an odds ratio ranging from 0.68 - 0.83 or 1.20 - 1.49. Estimates having no effect had an odds ratio ranging from 0.84 - 1.00 or 1.00 - 1.19. Model comparisons were further reviewed using receiver operating characteristic (ROC) methods, Z-scores, and p values (Hanley & McNeil, 1982). Correlation coefficients were used to examine hospital ranks (Alexandrescu, 2011).

Propensity Score Matching

Question 6 and Question 7 conditioned propensity score matching with stratification on the observed sample for the comparison of treatment effects in modeling the cesarean delivery outcome. The expectation was that there would be a difference in risk adjustment effects between the propensity score matched sample and the observational design sample. Second, the risk for a cesarean delivery among the matched sample would increase according to strata.

6. RQ 6: Is there a difference in cesarean delivery risk adjustment effects when comparing the propensity score matched sample versus the observational design sample?

$H_0$6: There is no difference in cesarean delivery risk adjustment effects when comparing the propensity score matched sample versus the observational design sample?

H_a6: There is a difference in cesarean delivery risk adjustment effects when comparing the propensity score matched sample versus the observational design sample?

7. RQ 7: For the propensity score matched sample, is there a difference in cesarean delivery risk adjustment effects by stratum?

$H_0$8: For the propensity score matched sample, there is no difference in cesarean delivery risk adjustment effects by stratum.

H_a7: For the propensity score matched sample, there is a difference in cesarean delivery risk adjustment effects by stratum.

SPSS Version 21.0, with the R plug-in for PS Matchit, was used for propensity score matching with stratification (Thoemmes, 2012). The 10 Robson Groups, 16 clinical conditions, and two SES independent variables were conditioned on the cesarean delivery outcome for the generation of a patient level predicted probability or the true propensity score. Once the propensity score was generated from the logistic regression model, one-to-one matching, and without replacement was conducted. This method assured cesarean delivery treatment cases were matched to a vaginal delivery control based on similarity in the propensity score. Caliper matching at 0.20 was selected and acted as the maximum distance between two events for match selection. This procedure produced a matched sample of cesarean delivery treatment cases, and vaginal controls with unmatched observations were discarded. The PS Matchit subclass feature was used to create 10 strata automatically based on the matched sample. The resulting data file was identified as the propensity score matched sample with stratification. To check the adequacy of the propensity score matched file, standardized mean differences of the cesarean delivery treatment cases and vaginal controls before and after matching were compared along with graphical diagnostics (Thoemmes, 2012).

To understand the stratum specific effects on the cesarean delivery outcome, a logistic regression model was created for each of the 10 subclasses. Similar to previous methods, variables for model entry were screened using chi-square test statistics with variables selection at the 0.05 level. As with previous approaches, the logistic regression models and their odds ratios were evaluated according to level of association (Handler, 1998) and ROC methods (Hanley & McNeil, 1982).

Threats to Validity

Threats to internal validity were concerned with data structure, risk adjustment, and statistical technique. Without external validation with other hospital informational sources, the 2012 linked file lacked representativeness of nonfederal births and clinical domains not collected through administrative data sources and associated with the cesarean procedure. Risk adjustment models, hospital ranks, and propensity score methods were at risk for selection bias, given some variables may be associated with the outcome but unavailable for modeling (Rosenbaum & Rubin, 1983). Statistical models may be biased due differences in approaches with HGLM modeling for multilevel data and logistic regression not having this capability (Huang, Dominici, Frangakis, et al, 2005). For this reason, comparative models were introduced to understand potential threats to internal validity by examining variation in risk adjustment, statistical technique, hospital ranks, and measures of association.

Study findings may not be generalizable posing a threat to external validity. The sampling frame was for the state of Georgia, and it may differ from other populations given variation in health care system structures, processes, and outcomes. Propensity score matching with stratification was introduced to reduce threats to external validity, given it models a randomized control trial as applied to the observational design. Still, these methodologies require further validation through statistical simulation studies.

Ethical Review

The research protocol received Institutional Review Board (IRB) approval through Walden University (IRB # 06-19-13-0124810). Data release was through the state of Georgia Department of Public Health, Office of Health Indicators and Planning (Appendix C). This research advanced an existing protocol between The Centers for Disease Control and Prevention and the state of Georgia (Giles et al., 2010). The state of Georgia required the research to mask hospital names and their volume deliveries due to the possibility of identifying facilities having small or large caseloads. Instead, the research protocol assigned a unique identification number to each hospital, and this was used for reporting.

Summary

This was a retrospective, observational design using a 2012 population-based linked data from the state of Georgia. Methodologies applied data linking, hospital profiling and propensity score matching with stratification. The cesarean delivery procedure was the study outcome and Robson 10 Group Classification, clinical, sociodemographic, and propensity score variables were used to create logistic regression and HGLM risk adjustment models for the generation of hospital ranks and measures of association. Propensity score matching with stratification was applied to the observational design for a review of model effects and measures of association. The next chapter reviews the findings from the research.

Chapter 4: Results

Introduction

The following chapter presents the findings of hospital profiling of the cesarean delivery procedure and propensity score matching with stratification.

Deterministic Data Linking

Deterministic data linking methods were required to prepare the linked maternal file using a 15-digit unique ID and matching algorithms including other variables.

Hospital Discharge Summary Data Extraction Methods

The 2012 hospital discharge summary file included approximately 1.1 million records from all nonfederal hospitals in the state of Georgia. Vaginal and cesarean delivery cases were first extracted from the 2012 hospital discharge summary file and other classifications of hospital admission were excluded from the analysis plan. Extraction types were defined according to method of delivery, live birth outcome, and parity outcomes including whether a singleton birth, multiple, or mixed with live birth and fetal death (Table 2). Extraction algorithms used Diagnosis Related Group (DRG), ICD-9-CM procedure, and ICD-9-CM diagnosis codes for the selection of vaginal and cesarean cases. The algorithm strategy first searched DRG codes sets because of an increase likelihood of case ascertainment with further iterations using ICD-9-CM procedure and diagnosis code sets. This extraction resulted in 127,414 unique maternal records from the 2012 hospital discharge summary file and eligible for the data linking procedure. When compared to the 2012 live birth file, including 130,661 unique records for linking, the hospital discharge summary record approximated 97.5% opportunities for a possible link.

Data Linking Performance

The linked file used in this research included 123,145 observations (Figure 1). The resulting file was derived by applying deterministic data linking algorithms to the 2012 Georgia live birth file (*N* = 130,661) and the 2012 Georgia hospital discharge summary file (*N* = 127,414). The first data linking algorithm used a 15-digit unique ID and matched events using the 2012 Georgia live birth file and the 2012 Georgia hospital discharge summary file. This pass resulted in 109,305 linked records, or an 88.7% data linking rate.

The second data linking algorithm used concatenation and manual review methods for the extraction of mother's date of birth from the 2012 hospital discharge summary file's 15-digit unique ID and the 2012 live birth file's 15-digit unique ID. The resulting extraction produced an 8-digit maternal date of birth for each file by record. The second pass linked the 2012 Georgia live birth file's 8-digit maternal date of birth, facility code, date of birth, event date, admission date and discharge date from each of the respective files. This resulted in an additional 10,136 linked records, and these were appended to the first data linking pass, resulting in 119,441 linked records, or 96.9% of the final linked file.

The remaining unlinked records included a nonstandard 14-digit unique ID. As with the second algorithm, the third data linking algorithm used concatenation and manual review methods to extract the mother's date of birth from the 14-digit ID by eliminating the first five characters and the last character from the alphanumeric string. The same data linking algorithm approach used in previous passes were applied and resulted in an additional 3,704 linked records, and these were appended to the 119,441 linked records, resulting in 124,135 cases. The final linked file represented 96.6% of the records from the original hospital discharge summary file and 94.2% of the live birth file. Approximately 5.8% of live birth and 3.4% of hospital discharge summary records remained unlinked.

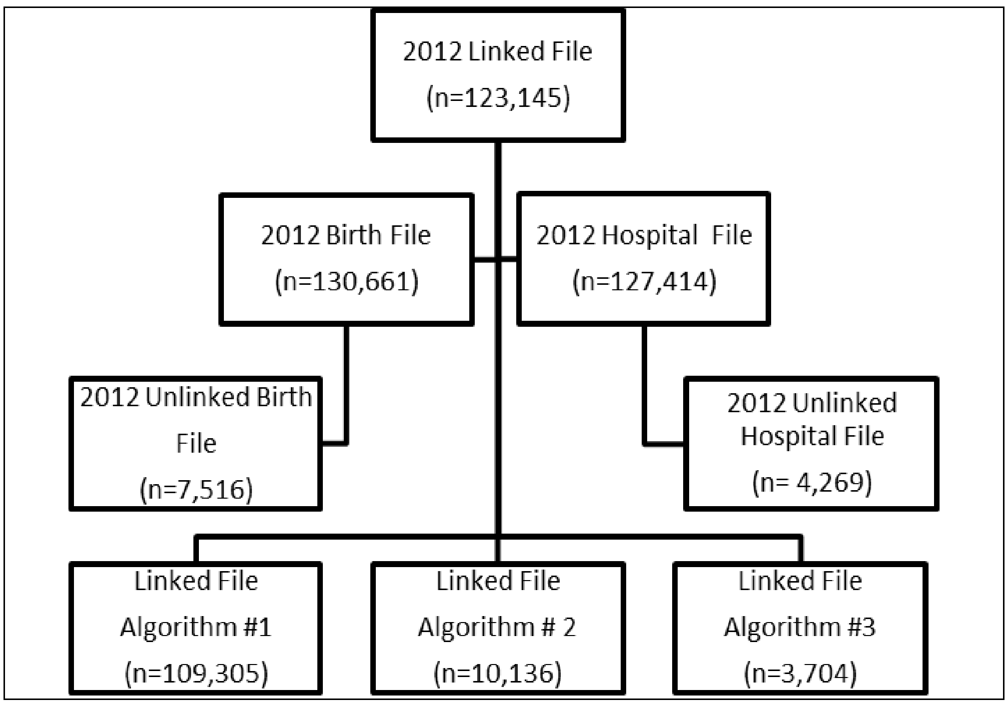

Figure 1. 2012 Georgia Maternal Linked Hospital Discharge Summary & Live Birth File

Findings demonstrated a difference in the performance of deterministic data linking algorithms by type. A data linking algorithm using the 15-digit ID highly performed, resulting in 88.7% records linked. Performance was improved to 96.6% when concatenation and manual review methods were introduced.

Clinical Descriptive Statistics

Sixteen clinical characteristics were extracted from the linked data set using ICD-9-CM principle diagnosis and live birth standard code sets. The 2012 live birth file did not include a number of expected variables, and this was due to under reporting by facilities. However, the hospital discharge summary file was used to extract the majority of maternal and infant risk factors for the creation of the analysis file.

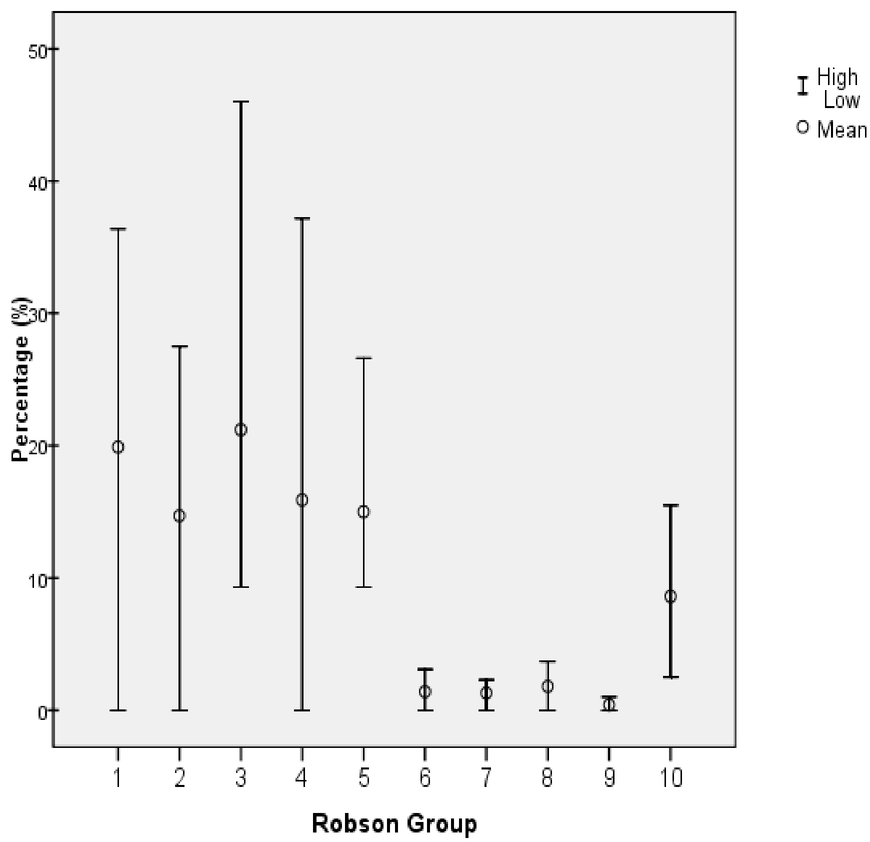

Figure 2. Hospital Distribution of Robson 10 Group Classification (*N* = 86)

For the period January 1, 2012 to December 31, 2012, the linked data file included 123,145 women with a live birth delivery. Among these, 43,231 had a cesarean section, and 79,914, a vaginal delivery. Among the 86 hospitals studied, the unadjusted cesarean section rate was 35.1% and ranged from 12.1% to 48.5% (Figure 3). Only one hospital had a cesarean delivery rate less than 20%, and more than 50 facilities had an unadjusted rate greater than 30%.

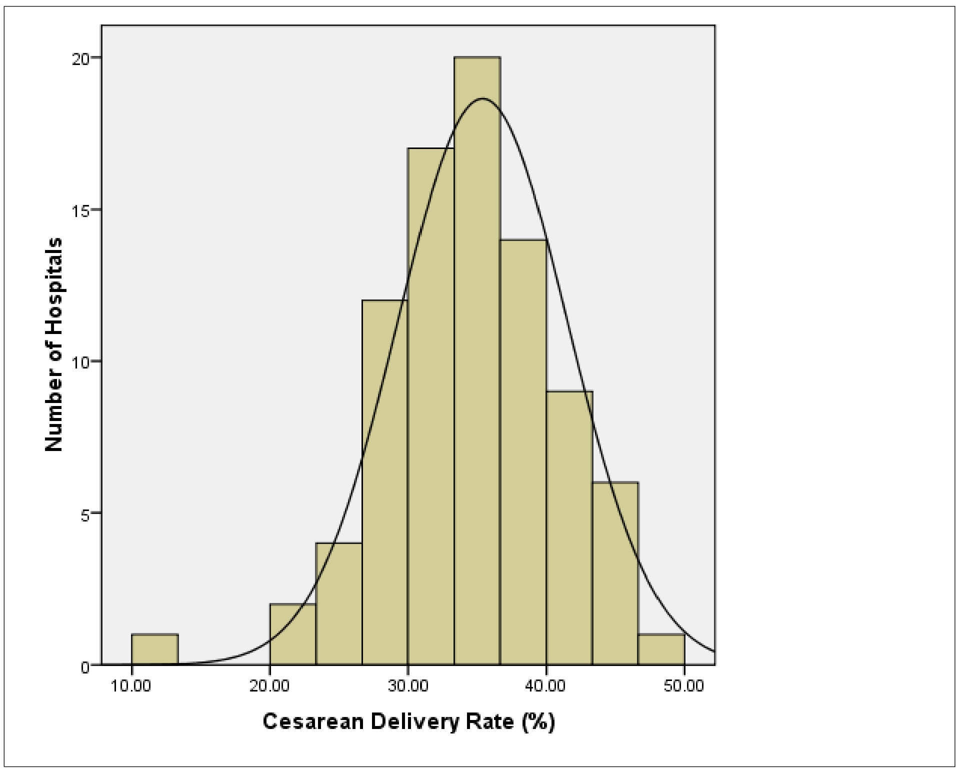

Figure 3. 2012 Georgia Cesarean Delivery Rate by Number of Hospitals (*N* = 86)

Among total live birth deliveries, 48.7% of women had no clinical risk factors indicated via hospital discharge summary ICD code sets, and 51.3% with one or more. Clinical characteristics significantly varied by rate, method of delivery, and at the facility level. Clinical conditions presented with the highest rates included dystocia (10.2%), fetal heart (13.9%) and anemia (14.0%), (Table 8). Rare events at the population level included fetal anomaly (1.7%), antepartum condition (1.7%), polyhydramnios (1.2%), oligohydramnios (2.8%), genital herpes (1.2%), and HIV (0.2%), (Table 8). Significant variation also existed when comparing clinical characteristics by method of delivery (Table 9). Anemia (18.6%), fetal heart (18.9%), dystocia (20.5%), fetal distress (9.9%), hypertension (9.9%), diabetes (8.7%), and eclampsia (8.1%) significantly differed when comparing cesarean and vaginal deliveries. For all clinical conditions, there was a significant difference between vaginal and cesarean deliveries, except for RH sensitization (Pearson Chi Square: $p < .503$) (Table 9).

Among the 86 hospitals, clinical conditions significantly varied (Figure 4). Anemia represented the broadest distribution and ranged from a high of 65.7% in one facility to a low of 0.4% in another (Pearson Chi Square: $p < .001$). Fetal heart had a rate of 31.8% in one facility and a low of 0.0% in another (Pearson Chi Square: $p < .001$). The rate of fetal distress significantly varied from 13.4% in one facility to 0.0% another (Pearson Chi Square: $p < .001$). Eclampsia also significantly varied among the 86 facilities, with a high of 15.2% to a low of 0.0% (Pearson Chi Square: $p < .001$). Oligohydramnios ranged from a high of 13.3% in one facility and 0.0% in another (Pearson Chi Square: $p < .001$). Lastly, hypertension ranged from a high of 16.2% to a low of 1.2% in another facility (Pearson Chi Square; $p < .001$).

Table 8

Maternal Clinical Condition and SES Descriptive Statistics (N = 123,145)

Variable	*N*	%
Clinical Condition		
Diabetes	7,466	6.1
Hypertension	9, 032	7.3
Eclampsia	5,783	4.7
HIV	238	0.2
Oligohydramnios	3,404	2.8
Polyhydramnios	1,487	1.2
Fetal abnormality	2,116	1.7
Antepartum	2,128	1.7
Fetal distress	5,592	4.5
Anemia	17,270	14.0
Dystocia	12,566	10.2
Fetal anomaly	2,089	1.7
RH	2,953	2.4
PROM	3,426	2.8
Genital herpes	1,482	1.2
Fetal heart	17,093	13.9
Insurance		
Public	68,510	55.6
Private	51,409	41.7
None	3,226	2.6
Maternal Age		
< 20	11,155	9.1
20 – 24	32,213	26.2
30 – 29	34,079	27.2
30 – 35	32,614	26.5
> 35	13,084	10.6

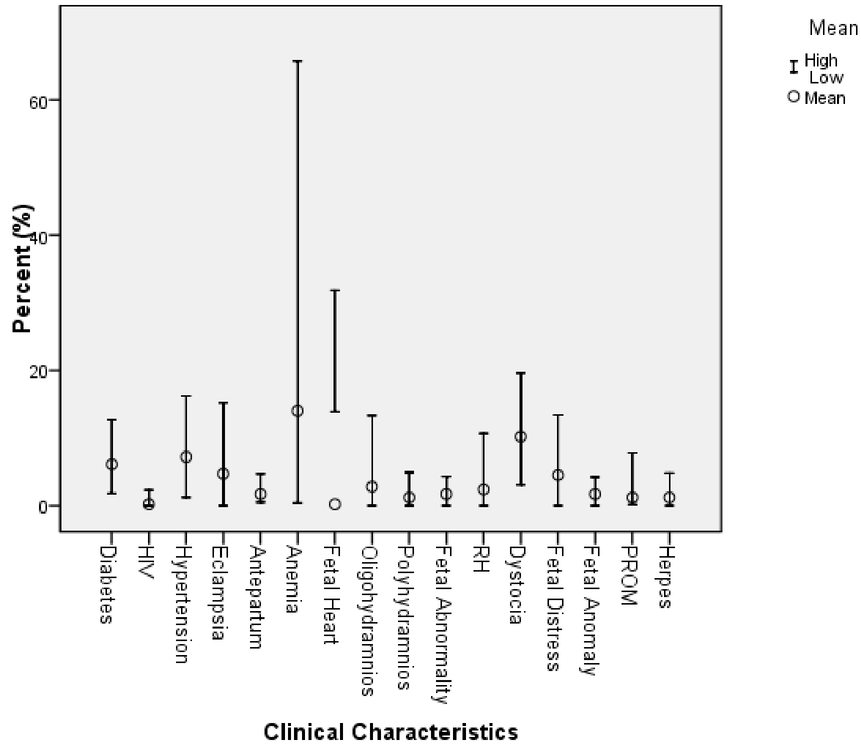

Figure 4. Distribution of Hospital Rates by Maternal Clinical Condition

(N = 86)

Table 9

Maternal Clinical Condition and SES Characteristics by Method of Delivery

(N = 123,145)

Variable	Cesarean delivery *N*	Cesarean delivery %	Vaginal delivery *N*	Vaginal delivery %	Pearson chi square X^2
Clinical condition					
Diabetes	3,768	8.7	3,718	4.7	< .001
Hypertension	4,269	9.9	4,763	6.0	< .001
Eclampsia	3,508	8.1	2,275	2.8	< .001
HIV	161	0.4	77	0.1	< .001
Oligohydramnios	1,607	3.7	1,797	2.2	< .001
Polyhydramnios	923	2.1	564	0.7	< .001
Fetal abnormality	1,025	2.4	1,091	1.7	< .001
Antepartum	1,439	3.3	689	0.9	< .001
Fetal distress	4,295	9.9	1,297	1.6	< .001
Anemia	8,042	18.6	9.228	11.5	< .001
Dystocia	8,874	20.5	3,682	4.6	< .001
Fetal anomaly	1,008	2.3	1,081	1.4	< .001
RH	1,037	2.4	1,916	2.4	< .503
PROM	1,302	3.0	2,124	2.7	< .001
Genital herpes	871	1.4	611	1.1	< .001
Fetal heart	8,160	18.9	8,933	11.2	< .001
Insurance					
Public	22,720	52.6	45,790	57.3	< .001
Private	19,520	45.2	31,889	39.9	< .001
None	991	2.3	2,235	2.8	< .001
Maternal age					
< 20	2,707	6.3	8,448	10.6	< .001
20 – 24	9,773	22.6	22,440	28.1	< .001
25 -29	11,590	26.8	22,489	28.1	< .001
30 – 35	12,879	29.8	19,735	24.7	< .001
> 35	6,282	14.5	6,802	8.5	< .001

Sociodemographic Characteristics

Maternal age varied, with the majority of live births occurring among women 20 to 35 years of age (Table 9). The majority of women with a live birth were publically insured (55.6%), followed by private insurance (41.7%) and having no insurance (2.6%) (Table 8). Having a vaginal delivery was more likely among the publically insured and those having no insurance ($p < .001$) (Table 9). A cesarean delivery was more likely among privately insured women ($p < .001$) (Table 10).

Robson 10 Group Descriptive Statistics

Research Question 1 tested whether there was a difference in the mean cesarean delivery rate among the Robson 10 Groups.

Findings from the extraction methodology demonstrated all maternal cases from the linked file were assigned to one of 10 Robson Groups with no missing events. Completeness was attributed to the availability of gestational age, parity, and plurality from the live birth file, which otherwise would not have been achieved by using the hospital discharge summary file alone. However, the hospital discharge summary file included concepts unique to its file that were not available from the live birth file, including breech, transverse/oblique lie, induction, and previous cesarean delivery. Tests of concordance and discordance were not achieved because the Robson indicators could not be extracted from both the Georgia hospital discharge file and Georgia live birth file. Instead, Robson indicators were extracted, using multiple variable algorithms as applied to the live birth file, hospital discharge summary file, or both.

Table 10

Robson 10 Group Classification Descriptive Statistics (N = 123,145)

	Cesarean delivery (A)	Live birth delivery (B)	CS rate	Relative size	Contribution
Robson Group	*N*	*N*	%	%	%
1	7,714	24,526	31.5	19.9	6.3
2	5,420	18,131	29.9	14.7	4.4
3	2,556	26,066	9.8	21.1	2.1
4	1,221	19,537	6.2	15.8	1.0
5	16,885	18,440	91.5	14.9	13.7
6	1,621	1,663	97.5	1.4	1.3
7	1,535	1,612	95.2	1.3	1.2
8	1,681	2,179	77.1	1.8	1.4
9	383	437	87.6	0.4	0.3
10	4,215	10,554	39.9	8.6	3.4
	43,231	123,145 (C)			

Note. CS rate = (A/B) x 100%, Relative size = (B/C) x 100%, Contribution = (A/C) x 100%.

Robson Groups varied according to cesarean section rate (CS), relative size, and contribution made by each group to the total cesarean delivery rate (Table 10). One-way ANOVA tests suggested the mean cesarean delivery rate between Robson 10 Groups were statistically significant, [F (9, 123,135) = 8,795, $p < .001$]. Tukey multiple comparison test suggested Groups 1, 2, 3, 4, 8, and 10 were significantly different from each of its comparisons with other Groups, and at the $p < .05$ level. For Groups 5, 6, 7, and 9 the majority of group comparisons were statistically significant at the $p < .05$ level. Yet, among the 90 different Robson 10 Group multiple comparisons, only Group 9 v. Group 5 ($p < .987$), Group 9 v. Group 7 ($p < .074$), and Group 7 v. Group 6 ($p < .779$) were not statistically different. Robson Group 5 (13.7%), Group 1 (6.3%) and Group 2

(4.4%) contributed the highest proportion to the overall cesarean delivery rate. Although smaller in size, Robson Group 6 (97.5%), Group 7 (95.2%), Group 8 (77.1%), and Group 9 (87.6%) represented high cesarean rates according to within group examinations. Group 5, or women with a previous cesarean section, had a within group cesarean section rate of 91.5% and reflected the national expectation.

Among the two groups classified as low risk pregnancies, within group comparisons demonstrated similarities among Group 1 cesarean deliveries with spontaneous labor (31.5%) and Group 2 having an induction procedure (29.9%). Among Robson Group 9, the within group expected number of cesarean deliveries was 100% versus an observed at 87.6%. Group 10 represented preterm births and had a within group cesarean section rate of 39.9% and contributed to the overall cesarean delivery rate by 3.4%.

Robson Groups significantly varied by facility (Pearson Chi-Square: $p < .001$) (Figure 2). Group 3 and Group 4 demonstrated a broad range in distribution by facility, with both having rates greater than 35%. However, Group 1 and Group 2 presented facilities with no cases (0.0%), with other facility rates exceeding 27.5% and 36.4%, respectively. Group 5 varied from a low of 9.3% in one facility to a high of 26.6% in another. Although smaller in size, Groups 6 to 9 demonstrated significant variation in hospital group as well. Among preterm deliveries, Group 10 had a low of 2.5% in one facility and a high of 15.5% in another.

Findings from the Robson extraction demonstrated near expected estimates for the majority of groups when compared to international standards. The relative size of Group

1 and Group 2 was approximately 33.6% when compared to an expected of rate of 35% to 42% (Betran et al., 2014). The combination of Group 2 and Group 4 was 37.1% as compared to an expected rate of 30% to 40% (Betran et al., 2014). Group 9 had a relative size of 0.4% versus an expected rate of 0.2% to 0.6%. Yet Group 9's observed cesarean delivery rate was 86% versus an expected 100% (Betran et al., 2014). This differential bias may be attributed to the quality of the hospital discharge summary data, standard code sets used, or abnormal lies classified as abnormal lies but delivered by vaginal delivery, such in the case of brow or face presentations. Women with a previous cesarean sections contributed 13.7% to the total cesarean rate, with the majority of women with a uterine scar having a cesarean delivery procedure (91.4%). This latter rate is congruent with population-based trends as measured by national indicators (NCHS, 2010).

Hospital Profiling

Research Questions 2 to 5 compared risk adjustment and statistical techniques for understanding the variation in cesarean delivery hospital ranks and model effects.

Statistical Comparative Models

Research Question 2 tested whether there was a difference in cesarean delivery risk adjustment effects when comparing logistic regression versus hierarchical generalized linear models.

When comparing logistic regression and hierarchical generalized linear models, findings varied according to risk adjustment method and hospital rankings (Table 11). Foremost, findings demonstrated the HGLM and logistic regression models did not differ by Robson 10 Group, clinical and SES model effects. Rather the findings for both models are presented as one table (Table 11). There were noted differences between model effects when comparing the Clinical Condition Model with the Full Model. For example, 10 of the 16 clinical conditions presented point estimates having non-overlapping confidence intervals suggesting a difference.

Findings from the Full Model, for Group 6, or women classified as nulliparous breech, showed an increase odds for a cesarean delivery (*OR* = 3.84, 95% CI [2.81, 5.26], $p < .001$) when compared Robson Group 5 or the referent group (Table 11). Groups 1 - 4 and Groups 8 - 10 had increased odds for a vaginal delivery. Among clinical conditions, women with a cesarean delivery had strong association with dystocia (*OR* = 13.31, 95% C.I. [12.62, 14.03], $p < .001$), fetal abnormality (*OR* = 3.49, 95% CI [1.79, 6.82], $p < .001$), HIV (*OR* = 10.37, 95% CI [7.27, 14.78], p < .001), eclampsia (*OR* = 4.34, 95% CI [4.03, 4.66], $p < .001$), fetal distress (*OR* = 16.09, 95% CI [14.81, 17.49], $p < .001$), antepartum (*OR* = 7.18, 95% CI [6.38, 8.07], $p < .001$), and fetal heart (*OR* = 4.45, 95% CI [4.24, 4.66], $p < .001$) when compared with those having a vaginal delivery. Other clinical conditions having a moderate association with the cesarean delivery outcome included genital herpes (*OR* = 2.27, 95% CI [1.94, 2.65], $p < .001$), oligohydramnios (*OR* = 2.06, 95% CI [1.87, 2.28], $p < .001$), polyhydramnios (*OR* = 2.35, 95% CI [2.01, 2.76], $p < .001$), anemia (*OR* = 1.95, 95% CI [1.85, 2.05], $p < .001$), diabetes (*OR* = 1.51, 95% CI [1.40, 1.62], $p < .001$), and hypertension (*OR* = 1.95, 95% CI [1.83, 2.08], $p < .001$). Among SES characteristics, women less than twenty years of age (y.o.a), 20 - 24 y.o.a. and 25 - 29 y.o.a. had an moderate association of having a vaginal delivery when compared to women 30 to 35 y.o.a.

Table 11

Risk Adjustment Effects on the Cesarean Delivery Outcome (N = 123,145)

Variable	Robson Model		Clinical Model		Full Model	
	OR	95% CI	*OR*	95% CI	*OR*	95% CI
Robson Group						
Group 5	Ref.				Ref.	
Group 1	0.04**	[0.03, 0.05]			0.02**	[0.01, 0.03]
Group 2	0.04**	[0.03, 0.05]			0.02**	[0.01, 0.03]
Group 3	0.02**	[0.01, 0.03]			0.02**	[0.02, 0.03]
Group 4	0.02**	[0.01, 0.03]			0.02**	[0.01, 0.03]
Group 6	3.67**	[2.67, 5.01]			3.84**	[2.81, 5.26]
Group 7	1.92**	[1.52, 2.43]			1.64**	[1.29, 2.08]
Group 8	0.31**	[0.26, 0.35]			0.22**	[0.19, 0.25]
Group 9	0.80	[0.60, 1.06]			0.46**	[0.34, 0.62]
Group 10	0.06**	[0.05, 0.07]			0.03**	[0.02, 0.04]
Clinical Condition						
Diabetes			1.67**	[1.59, 1.76]	1.51**	[1.40, 1.62]
Hypertension			1.62**	[1.55, 1.70]	1.95**	[1.83, 2.08]
Eclampsia			3.05**	[2.88, 3.23]	4.34**	[4.03, 4.66]
HIV			5.15**	[3.89, 6.84]	10.37**	[7.27, 14.78]
Oligohydramnios			1.62**	[1.50, 1.74]	2.06**	[1.87, 2.28]
Polyhydramnios			2.20**	[1.96, 2.47]	2.35**	[2.01, 2.76]
Fetal abnormality			1.93*	[1.18, 3.19]	3.49**	[1.79, 6.82]
Antepartum			4.32**	[3.93, 4.76]	7.18**	[6.38, 8.07]
Fetal distress			5.35**	[5.01, 5.73]	16.09**	[14.81, 17.49]
Anemia			1.82**	[1.76, 1.80]	1.95**	[1.85, 2.05]
Dystocia			4.66**	[4.47, 4.87]	13.31**	[12.62, 14.03]
Fetal anomaly			0.82	[0.49, 1.35]	0.59	[0.31, 1.17]
RH			1.01	[0.92, 1.09]	1.02	[0.91, 1.15]
PROM			1.18**	[1.09, 1.29]	0.99#	[0.81, 0.99]
Genital herpes			1.36**	[1.23, 1.52]	2.27**	[1.94, 2.65]
Fetal heart			1.61**	[1.56, 1.67]	4.45**	[4.24, 4.66]
Insurance						
Public					Ref.	
Private					0.91**	[0.88, 0.95]
None					0.76**	[0.67, 0.86]
Maternal age						
30 - 35					Ref.	
< 20					0.45**	[0.41, 0.48]
20 - 24					0.62**	[0.59, 0.65]
25 - 29					0.45**	[0.73, 0.81]
> 35					1.45**	[1.36, 1.55]

Note. # = p < .05, * = p < .01, ** = p <0.001, Ref. = referent category, OR = odds ratio, CI = confidence interval.

Propensity Score Risk Adjustment

Research Question 3 tested whether there was a difference in cesarean delivery risk adjustment effects when comparing logistic regression propensity score versus hierarchical generalized linear propensity score models.

Propensity score risk adjustment involved the examination of 4 models: (1) Logistic regression full model with the true propensity score added as a risk adjustor (Table 12); (2) Logistic regression full model with the transformed propensity score added as a risk adjustor (Table 12); (3) HGLM full model with the true propensity score added as a risk adjustor (Table 13); and (4) HGLM full model with the transformed propensity score added as a risk adjustor (Table 13). The Full Model included 10 Robson Groups, 16 clinical conditions and 2 clinical characteristics as risk adjustors and in modelling the cesarean delivery outcome.

Logistic regression was used to generate a propensity score for each patient using the 10 Robson Groups, 16 clinical conditions and two SES characteristics. Among the 123,145 maternity patients, the mean propensity score was 0.351 with a minimum of 0.00624 and a maximum of 1.0000. The distribution of the true propensity score was a continuous variable (Figure 5), and log transformation was applied to each patient’s score in an attempt to produce a normal distribution (Figure 6).

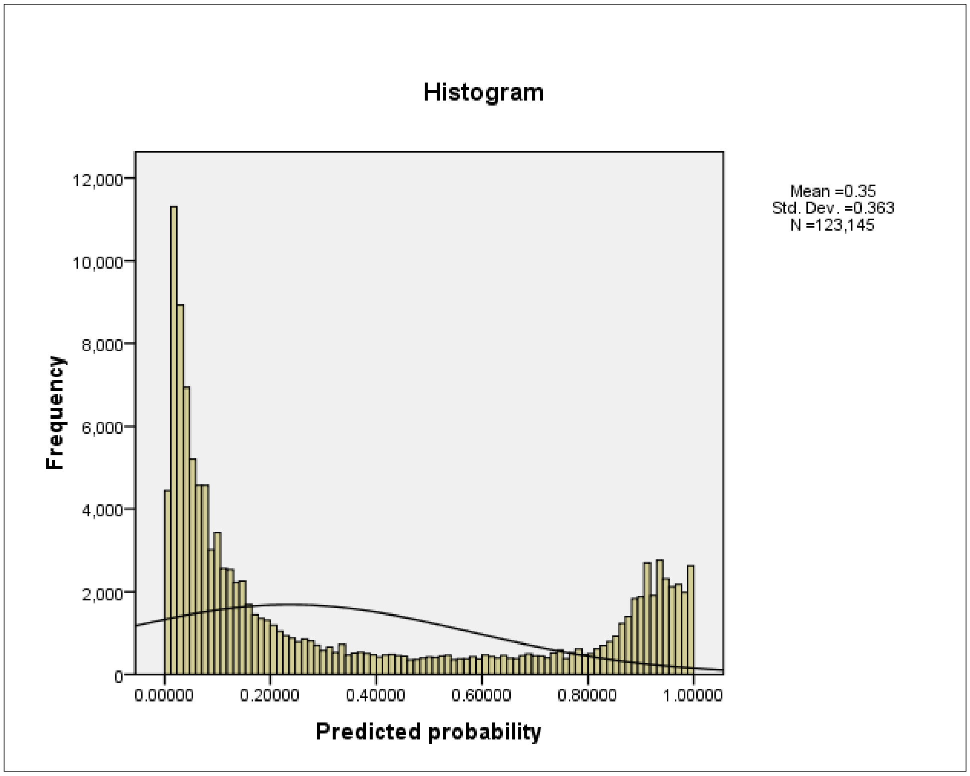

Figure 5. Frequency Count of Maternal Propensity Score (N = 123,145)

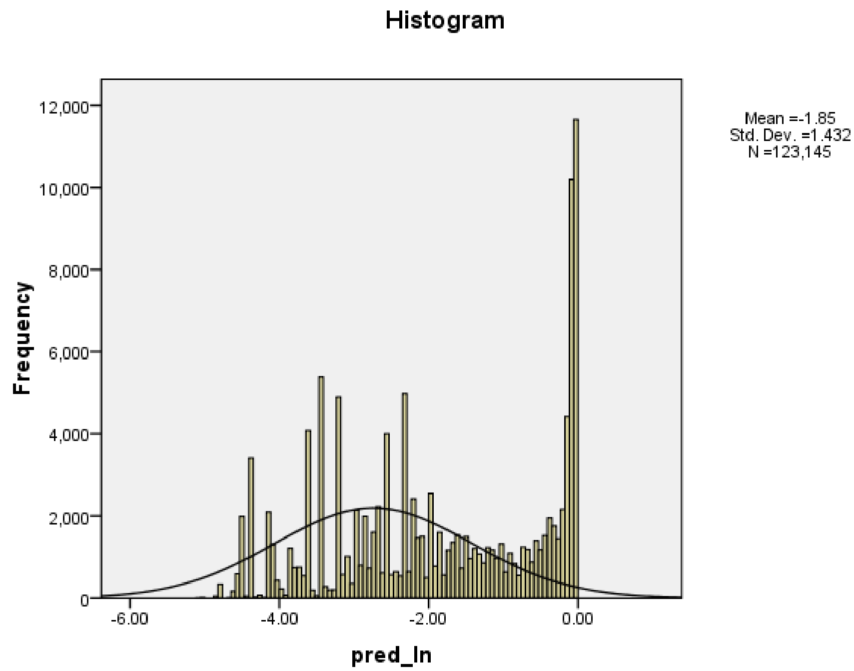

Figure 6. Frequency Count of Transformed Propensity Score (*N* = 123,145)

Table 12

Logistic Regression Propensity Score Risk Adjustment Effects on the Cesarean Delivery Outcome (N = 123,145)

Variable	True propensity score		Transformed propensity score	
	OR	95% CI	*OR*	95% CI
Robson Group				
Group 5	Ref.		Ref.	
Group 1	0.05^{**}	[0.04, 0.06]	0.12^{**}	[0.10, 0.13]
Group 2	0.03^{**}	[0.02, 0.05]	0.09^{**}	[0.08, 0.11]
Group 3	0.02^{**}	[0.01, 0.03]	0.08^{**}	[0.06, 0.09]
Group 4	0.02^{**}	[0.01, 0.03]	0.06^{**}	[0.04, 0.07]
Group 6	3.34^{**}	[2.43, 4.57]	3.31^{**}	[2.42, 4.52]
Group 7	1.56^{**}	[1.23, 1.98]	1.64^{**}	[1.29, 2.08]
Group 8	0.31^{**}	[0.27, 0.36]	0.35^{**}	[0.31, 0.39]
Group 9	0.54^{**}	[0.40, 0.73]	0.64^{*}	[0.47, 0.84]
Group 10	0.07^{**}	[0.05, 0.08]	0.13^{**}	[0.12, 0.15]
Clinical Condition				
Diabetes	1.38^{**}	[1.28, 1.49]	1.17^{**}	[1.09, 1.26]
Hypertension	1.68^{**}	[1.56, 1.80]	1.29^{**}	[1.21, 1.38]
Eclampsia	3.02^{**}	[2.71, 3.36]	1.79^{**}	[1.64, 1.95]
HIV	5.89^{**}	[4.03, 8.59]	2.42^{**}	[1.72, 3.42]
Oligohydramnios	1.74^{**}	[1.56, 1.94]	1.32^{**}	[1.19, 1.46]
Polyhydramnios	1.93^{**}	[1.64, 2.27]	1.44^{**}	[1.24, 1.68]
Fetal abnormality	2.69^{*}	[1.37, 5.26]	1.62	[0.85, 3.09]
Antepartum	4.35^{**}	[3.71, 5.09]	2.19^{**}	[1.93, 2.49]
Fetal distress	8.11^{**}	[6.86, 9.59]	3.20^{**}	[2.85, 3.59]
Anemia	1.68^{**}	[1.58, 1.79]	1.31^{**}	[1.24, 1.38]
Dystocia	6.88^{**}	[5.93, 7.99]	2.79^{**}	[2.52, 3.08]
Fetal anomaly	0.65	[0.33, 1.28]	0.82	[0.43, 1.57]
RH	1.11	[0.90, 1.15]	1.01	[0.89, 1.13]
PROM	0.93	[0.84, 1.02]	0.95	[0.86, 1.05]
Genital herpes	1.89^{**}	[1.61, 2.23]	1.36^{**}	[1.16, 1.58]
Fetal heart	3.15	[2.89, 3.44]	1.69^{**}	[1.57, 1.81]
Insurance				
Public	Ref.		Ref.	
Private	0.93^{**}	[0.89, 0.97]	0.97	[0.93, 1.01]
None	0.81^{**}	[0.71, 0.91]	0.89	[0.79, 1.02]
Maternal age				
30 - 35	Ref.		Ref.	
< 20	0.53^{**}	[0.49, 0.58]	0.74^{**}	[0.68, 0.82]
20 - 24	0.68^{**}	[0.64, 0.73]	0.83^{**}	[0.78, 0.87]
25 - 29	0.81^{**}	[0.77, 0.85]	0.89^{**}	[0.85, 0.95]
>35	1.34^{**}	[1.25, 1.43]	1.17^{**}	[1.09, 1.24]

Note. * = p < .01, ** = p < .001, Ref. = referent category, OR = odds ratio, 95% CI = 95% confidence interval.

Table 12 compared logistic regression models with the true propensity score (LR PS) added as a risk adjustor to the Full Model versus the transformed propensity score. Significant differences existed in the comparison of especially clinical conditions. For the logistic regression model including the true propensity score as a risk adjustor, women with eclampsia (*OR* = 3.02, 95% CI [2.71, 3.36], $p < .001$), HIV (*OR* = 5.89, 95% CI [4.03, 8.59], $p < .001$), antepartum (*OR* = 4.35, 95% CI [3.71, 5.09], $p < .001$), fetal distress (*OR* = 8.11, 95% CI [6.86, 9.59], $p < .001$), dystocia (*OR* = 6.88, 95% CI [5.93, 7.99], $p < .001$), and fetal heart (*OR* = 3.15, 95% CI [2.89, 3.44], $p < .001$) had a strong association with a cesarean delivery when compared to women having a vaginal delivery. Yet, for the logistic regression model with the transformed propensity score as a risk adjustor, only fetal distress (*OR* = 3.20, 95% CI [2.85, 3.59], $p < .001$) had a strong association with the cesarean delivery procedure.

Among clinical conditions having a moderate association with the cesarean delivery outcome, these also differed by risk adjustment model (Table 12). For the risk adjustment model with the true propensity score, hypertension (*OR* = 1.68, 95% CI [1.56, 1.80], $p < .001$), oligohydramnios (*OR* = 1.74, 95% CI [1.56, 1.94], $p < .001$), polyhydramnios (*OR* = 1.93, 95% CI [1.64, 2.27], $p < .001$), anemia (*OR* = 1.68, 95% CI [1.58, 1.79], $p < .001$), and genital herpes (*OR* = 1.89, 95% CI [1.61, 2.23], $p < .001$) had a moderate association with the cesarean delivery outcome. Yet, in the review of the risk adjustment model with the transformed propensity score added, these conditions demonstrated a weak association with the cesarean delivery procedure when compared to a vaginal delivery.

For HGLM models with the true propensity score and transformed propensity score added as risk adjustors, no clinical condition presented with a strong association with the cesarean delivery outcome, and this differed from the logistic regression propensity score risk adjustment models (Table 12) (Table 13). Rather, the HGLM model with the true propensity score added as a risk adjustor identified eclampsia (*OR* = 1.79, 95% CI [1.64, 1.97], $p < .001$), HIV (*OR* = 2.85, 95% CI [1.96, 4.15], $p < .001$), fetal abnormality (*OR* = 1.94, 95% CI [0.98, 3.83]), antepartum (*OR* = 2.02, 95% CI [1.75, 2.34], $p < .001$), fetal distress (*OR* = 2.82, 95% CI [2.48, 3.23], $p < .001$), dystocia (*OR* = 2.52, 95% CI [2.25, 2.83], $p < .001$), genital herpes (*OR* = 1.54, 95% CI [1.31, 1.81], $p < .001$), and fetal heart (*OR* = 1.97, 95% CI [1.85, 2.12], $p < .001$) as moderately associated with the cesarean delivery outcome when compared to vaginal deliveries (Table 13). For the HGLM model with the transformed propensity score added as a risk adjustor, dystocia (*OR* = 1.73, 95% CI [1.59, 1.89], $p < .001$), HIV (*OR* = 1.83, 95% CI [1.31, 2.56], $p < .001$), antepartum (*OR* = 1.53, 95% CI [1.36, 1.74], $p < .001$), and fetal distress (*OR* = 1.93, 95% CI [1.74, 2.14]) demonstrated a moderate association with the cesarean delivery procedure. All other clinical conditions demonstrated a weak association or no effect.

For both HGLM models with the propensity score added as a risk adjustor, Group 1, Group 2, Group 3, Group 4, and Group 10 had a strong association with a vaginal delivery in comparison to the Group 5 referent group. Only Group 7 had a moderate association with the cesarean delivery outcome when compared to Group 5 as the referent group.

Table 13

HGLM Propensity Score Risk Adjustment Effects on the Cesarean Delivery Outcome (N = 123,145)

Variable	True propensity score		Transformed propensity score	
	OR	95% CI	*OR*	95% CI
Robson Group				
Group 5	Ref.		Ref.	
Group 1	0.25**	[0.21, 0.29]	0.20**	[0.18, 0.23]
Group 2	0.19**	[0.16, 0.23]	0.17**	[0.16, 0.20]
Group 3	0.13**	[0.10, 0.16]	0.18**	[0.16, 0.22]
Group 4	0.07**	[0.05, 0.09]	0.15**	[0.13, 0.19]
Group 6	2.86**	[2.09, 3.93]	3.37**	[2.47, 4.61]
Group 7	1.50**	[1.18, 1.90]	1.72**	[1.36, 2.18]
Group 8	0.52**	[0.45, 0.59]	0.39**	[0.35, 0.45]
Group 9	0.71*	[0.53, 0.95]	0.71*	[0.53, 0.95]
Group 10	0.29**	[0.25, 0.34]	0.20**	[0.18, 0.23]
Clinical Conditions				
Diabetes	1.20**	[1.11, 1.29]	1.09*	[1.02, 1.17]
Hypertension	1.33**	[1.24, 1.42]	1.12**	[1.05, 1.19]
Eclampsia	1.79**	[1.64, 1.97]	1.42**	[1.31, 1.54]
HIV	2.85**	[1.96, 4.15]	1.83**	[1.31, 2.56]
Oligohydramnios	1.38**	[1.24, 1.54]	1.17**	[1.06, 1.29]
Polyhydramnios	1.47**	[1.25, 1.74]	1.30	[1.12, 1.51]
Fetal abnormality	1.94	[0.98, 3.83]	1.44	[0.75, 2.75]
Antepartum	2.02**	[1.75, 2.34]	1.53**	[1.36, 1.74]
Fetal distress	2.82**	[2.48, 3.23]	1.93	[1.74, 2.14]
Anemia	1.45**	[1.37, 1.53]	1.27**	[1.20, 1.34]
Dystocia	2.52**	[2.25, 2.83]	1.73**	[1.59, 1.89]
Fetal anomaly	0.71	[0.36, 1.40]	0.81	[0.42, 1.55]
RH	1.03	[0.92, 1.17]	1.04	[0.93, 1.18]
PROM	0.97	[0.89, 1.08]	0.98	[0.88, 1.08]
Genital herpes	1.54**	[1.31, 1.81]	1.25**	[1.07, 1.46]
Fetal heart	1.97**	[1.85, 2.12]	1.32**	[1.24, 1.40]
Insurance				
Public	Ref.		Ref.	
Private	0.92**	[0.88, 0.96]	0.93	[0.89, 0.97]
None	0.86*	[0.76, 0.97]	0.92	[0.81, 1.04]
Maternal age				
30-35	Ref.		Ref.	
< 20	0.67	[0.62, 0.72]	0.81	[0.75, 0.87]
20 - 24	1.19	[1.12, 1.28]	0.87	[0.82, 0.92]
25 - 29	0.87	[0.82, 0.91]	0.93	[0.89, 0.98]
> 35	1.19	[1.12, 1.28]	1.09	[1.03, 1.17]

Note. # = p < .05, * = p < .01, ** = p < .001, Ref. = referent category, OR = odds ratio, CI = confidence interval.

Hierarchical generalized linear models with the propensity scores added as risk adjustor (Table 13) differed from the logistic regression models (Table 12). The HGLM models presented different determinants of association and narrower confidence intervals when compared to the logistic regression models. Determinants strongly associated with the cesarean delivery outcome as measured by the logistic regression models no longer had this magnitude of association, as demonstrated by the HGLM statistics.

When comparing the HGLM and logistic regression true propensity score models, 13 of 31 (42%) standardized model effects had non-overlapping confidence intervals suggesting a difference. Likewise, when comparing the HGLM and logistic regression transformed propensity score models, 9 of 31 (29%) model effects had non-overlapping confidence intervals suggesting a difference. For the logistic regression true propensity score model, 26 of 31 (84%) standardized model effects were statistically significant at $p < .001$. For the HGLM model, 19 of 31 (61%) model effects were statistically significant at $p < .001$. When comparing the HLGM and logistic regression p values for the transformed propensity score, the HGLM model indicated 17 of 31 (54%) effects were statistically significant at $p < .001$ versus the logistic regression model with 26 of 31 effects (84%).

Cesarean Delivery Hospital Ranks

Research Question 4 tested whether there was a difference in cesarean delivery hospital ranks when comparing logistic regression versus hierarchical generalized linear models. Whereas, Research Question 5 tested whether there was a difference in cesarean delivery hospital ranks when comparing propensity score risk adjustment models using logistic regression or hierarchical generalized linear models.

Table 14

Correlation Matrix of Hospital Ranks by Model Type (N = 86)

Model	LR Full	LR PS	LR LN	LR Null	HGLM Full	HGLM PS	HGLM LN	HGLM Null
LR Full	1.0	0.99*	0.99*	0.67*	0.42	0.45	0.41	0.05
LR PS	0.99*	1.0	0.98*	0.65*	0.41	0.44	0.40	0.06
LR LN	0.99*	0.98*	1.0	0.63*	0.41	0.44	0.40	0.05
LR Null	0.67*	0.65*	0.63*	1.0	0.67	0.66	0.66	0.04
HGLM Full	0.42	0.41	0.41	0.67	1.0	0.99*	0.99*	0.05
HGLM PS	0.45	0.44	0.44	0.66	0.99*	1.0	0.99*	0.03
HGLM LN	0.41	0.40	0.40	0.66	0.99*	0.99*	1.0	0.03
HGLM Null	0.05	0.06	0.05	0.04	0.05	0.03	0.03	1.0

Note. * = $p < .01$ (2-tailed), Spearman Correlation.

Hospital ranks varied according to risk adjustment model and statistical technique (Table 14). Hospital ranks derived from the logistic regression full model (LR Full) were highly correlated with ones developed from the logistic regression full model having the true propensity score added as a risk adjustor (LR PS) ($p < .01$). Similarly, hospital ranks derived from the logistic regression full model with the true propensity (LR PS) score were highly correlated with the ones derived from the logistic regression full model with the transformed propensity score (LR LN) ($p < .01$). Among HGLM models, similar patterns were noted: ranks derived from the HGLM full model (HGLM Full) were highly correlated with the HGLM full model with the true propensity score (HGLM PS) as a risk adjustor ($p < .01$). Likewise, ranks developed from the HGLM full model with the true propensity score (HGLM PS) were highly correlated with the HGLM full model with the transformed propensity score (HGLM PS) ($p < .01$).

Hospital ranks derived from logistic regression models were not correlated with the ones developed by the HGLM models. Notably, hospital ranks developed from the logistic regression full model (LR Full) and HGLM full model (HGLM Full), including ones with the true propensity score, or the transformed score, were not correlated with their respective null models. This finding emphasized the importance of hospital risk adjustment versus ranks typically generated by unadjusted rates.

Figure 7 used boxplots and compares eight differing risk adjustment and statistical technique models in the examination of the hospital log odds. For six of the models, whether logistic regression or HGLM, Hospital 47 was an outlier and represented one of the highest likelihoods of having a cesarean delivery when compared to the referent hospital. Among the three HGLM models with risk adjustment, Hospitals 12, 23, 25, 37, 40, and 47, were outliers, suggesting an increase in the odds of a cesarean delivery when compared to the referent hospital. These differed from the logs odds generated from the logistic regression models, with only Hospital 47 representing an outlier among risk adjustment models and representing an increase risk of cesarean delivery when compared to the referent hospital.

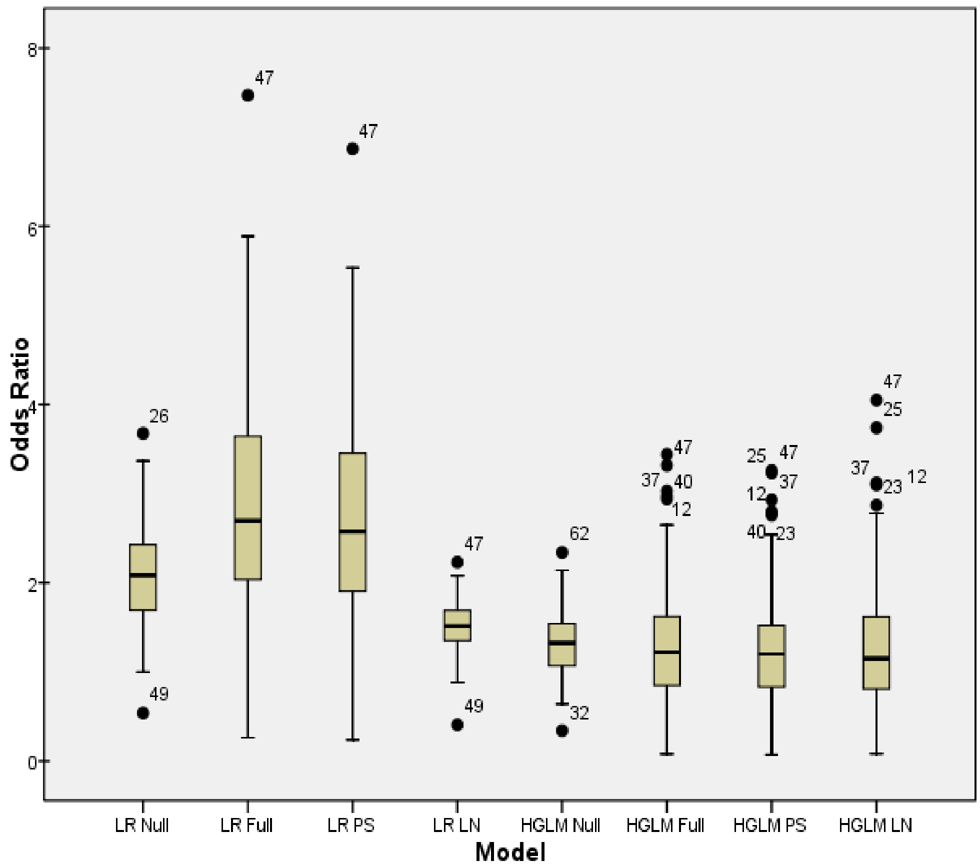

Figure 7. Boxplot Comparison of Hospital Odds Ratio by Cesarean Delivery Model Type (N = 86)

The boxplots graphically suggest the medium odds ratio and their quartile metrics varied by risk adjustment model and statistical technique. For example, the logistic regression models presented a wide variation in hospital log odds when compared to HGLM risk adjustment models. This suggested shrinkage estimates among the HGLM models when compared to the logistic regression models. Null models demonstrated the most conservative odds ratios and emphasized the importance of risk adjustment models and their variation.

Propensity Score Matching

Research Questions 6 and 7 applied propensity score matching with stratification to the observational design for the creation of a matched sample of cesarean delivery cases (treatment) and vaginal delivery controls.

Propensity Score Matching with Stratification

Research Question 6 tested whether there was a difference in cesarean delivery risk adjustment effects when comparing the propensity score matched sample versus the observational design sample. Propensity score matching with stratification was used to analyze the treatment and control groups for conditioning on the cesarean delivery outcome. A logistic regression model, with cesarean delivery as the outcome, used 10 Robson Groups, 16 clinical conditions and two sociodemographic variables as risk adjustors to generate the propensity score for each maternity patient in preparation for matching and stratification. One-to-one nearest neighbor matching, without replacement, at 0.20 calipers, and with stratification by 10 classes resulted in a sample of 33,820 observations, or 16,910 cesarean treatment cases and 16, 910 vaginal controls. From the original observed set of 123,145 cases, the number of unmatched cases included 26,321 women with a cesarean procedure and 63,004 having a vaginal delivery.

The Hansen and Bowers overall balance test was significant at $p < .001$. The multivariate imbalance before matching was 0.768 and 0.474 after matching. In the final matched model, no covariate exhibited a large imbalance of greater than 0.25 and no imbalances of observed covariates. Figures 8 to 12 depict the diagnostics used to determine whether the resulting matched sample met an expected covariate balance. Figure 9 demonstrates improved covariate balances with a change in the standardized difference for the propensity score for Robson Group 3, Robson Group 4, Robson Group 5, dystocia, and maternal age less than 20 years of age. Figure 10 also demonstrates improved balance based on standardized differences when comparing the matched data (n = 33,820) with the original observed data (N = 123,145). Likewise, Figure 12 demonstrates an improvement in standardized differences when comparing the prematched data (N = 123,145) with the matched data set (n = 33,820). Both Figures 8 and 11 demonstrate observations in the extreme tail areas and represented a proportion of unmatched observations.

Baseline differences between the unmatched cesarean treatment group and the unmatched vaginal control group existed and was measured by the true propensity score. Before matching, the probability of a cesarean delivery was higher among the cesarean treatment cases when compared to the vaginal delivery controls. The mean propensity score at baseline for the 79,914 vaginal delivery controls was 0.15 (SD = 0.20), whereas the propensity score at baseline for the 43,231 cesarean delivery cases had a mean propensity score of 0.726 (SD = 0.29). After matching, the two groups where comparable with the mean propensity score for the matched controls at 0.42 (SD = 0.01) and 0.47 (SD = 0.01) for the matched treated cases.

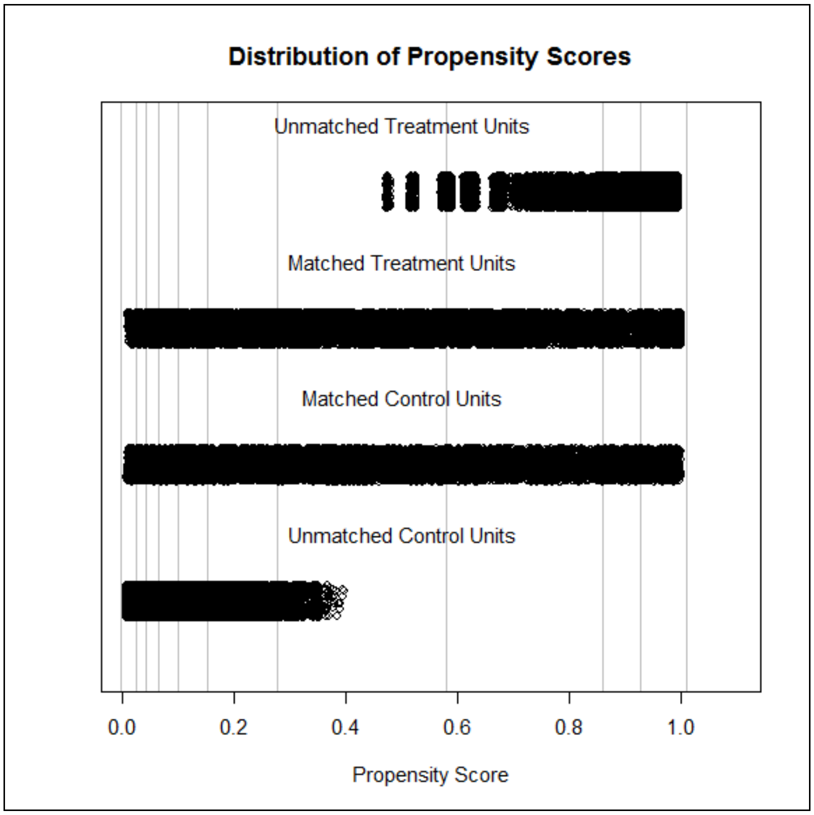

Figure 8. Propensity Score Distribution for Cesarean Treatment and Vaginal Controls (n = 33,820)

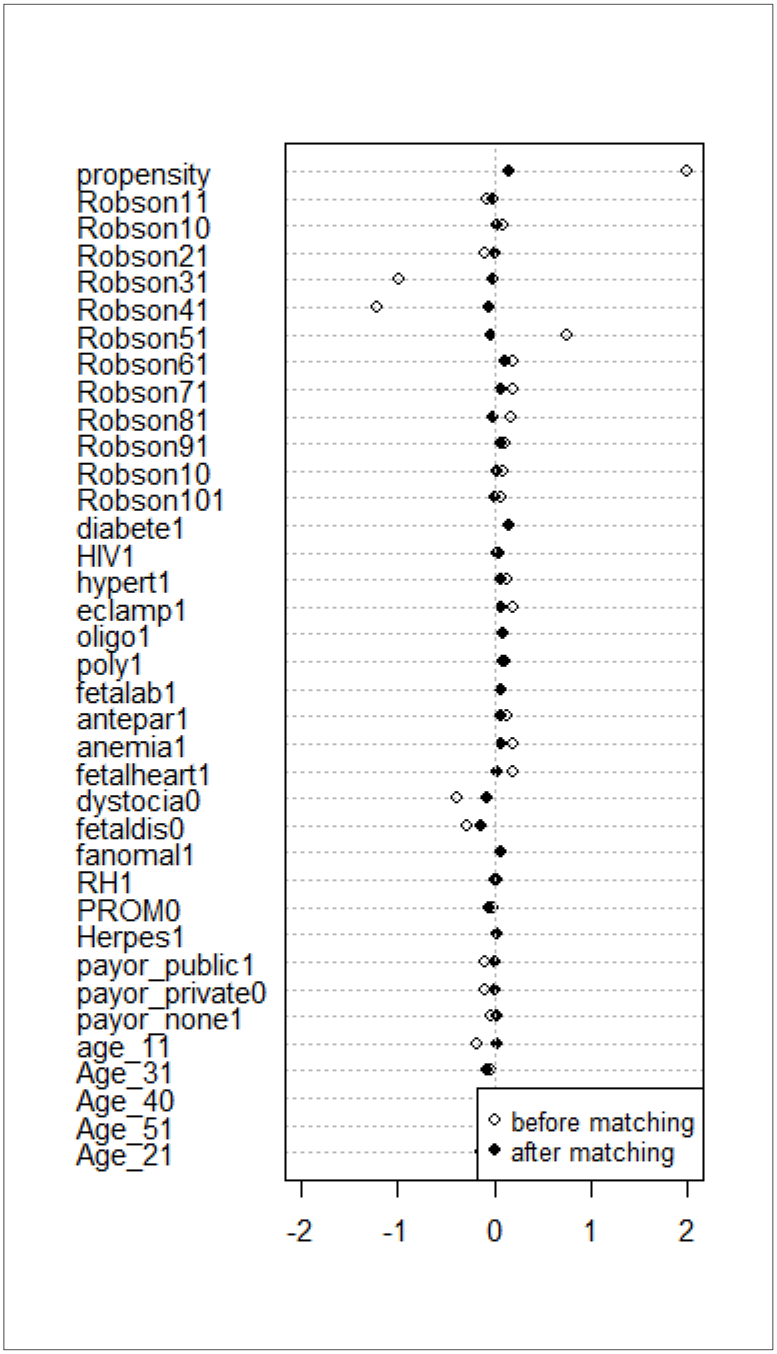

Figure 9. Comparative Kernel Estimates of Standardized Differences

(n = 33,820)

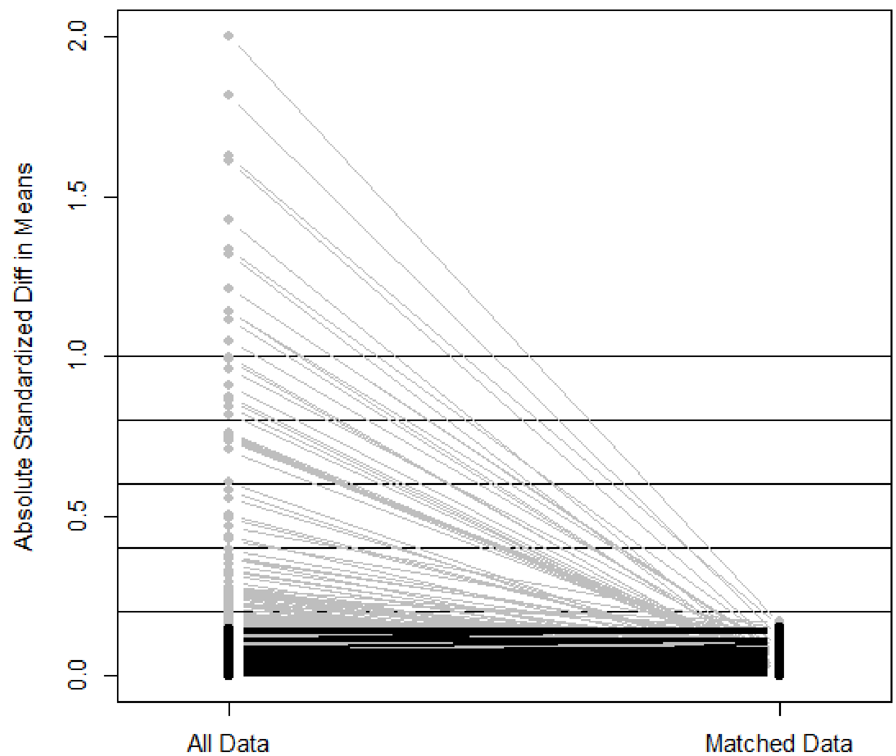

Figure 10. Standardized Differences in Means for Propensity Score Matched and Unmatched Data (n = 33,820)

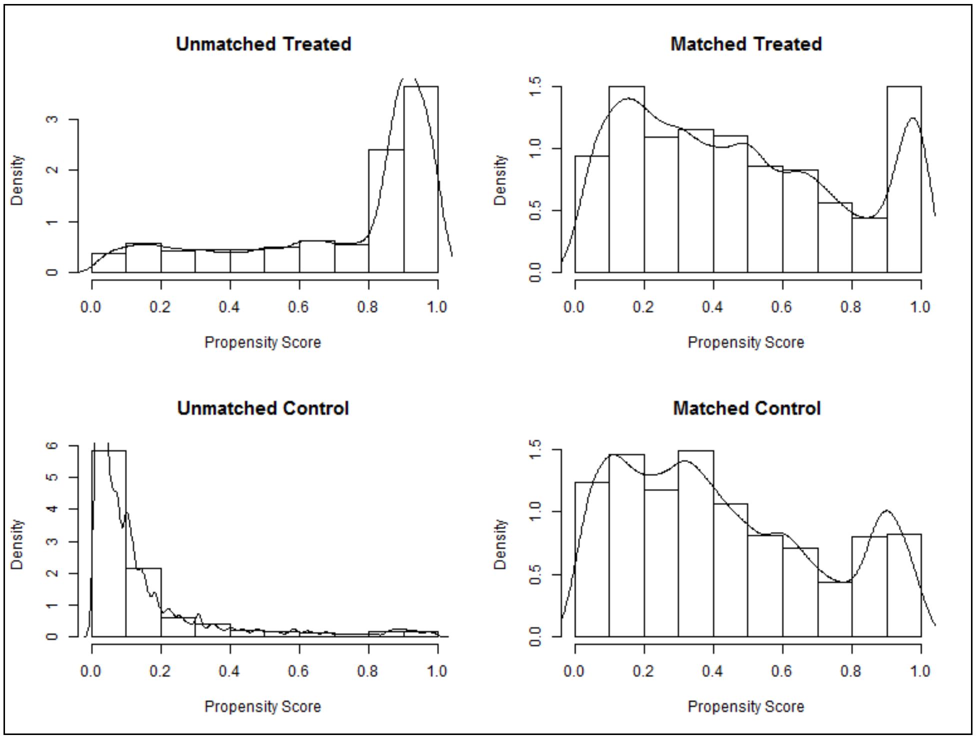

Figure 11. Propensity Score Density by Matched and Unmatched Cesarean Treatment and Vaginal Delivery Controls (n = 33,820)

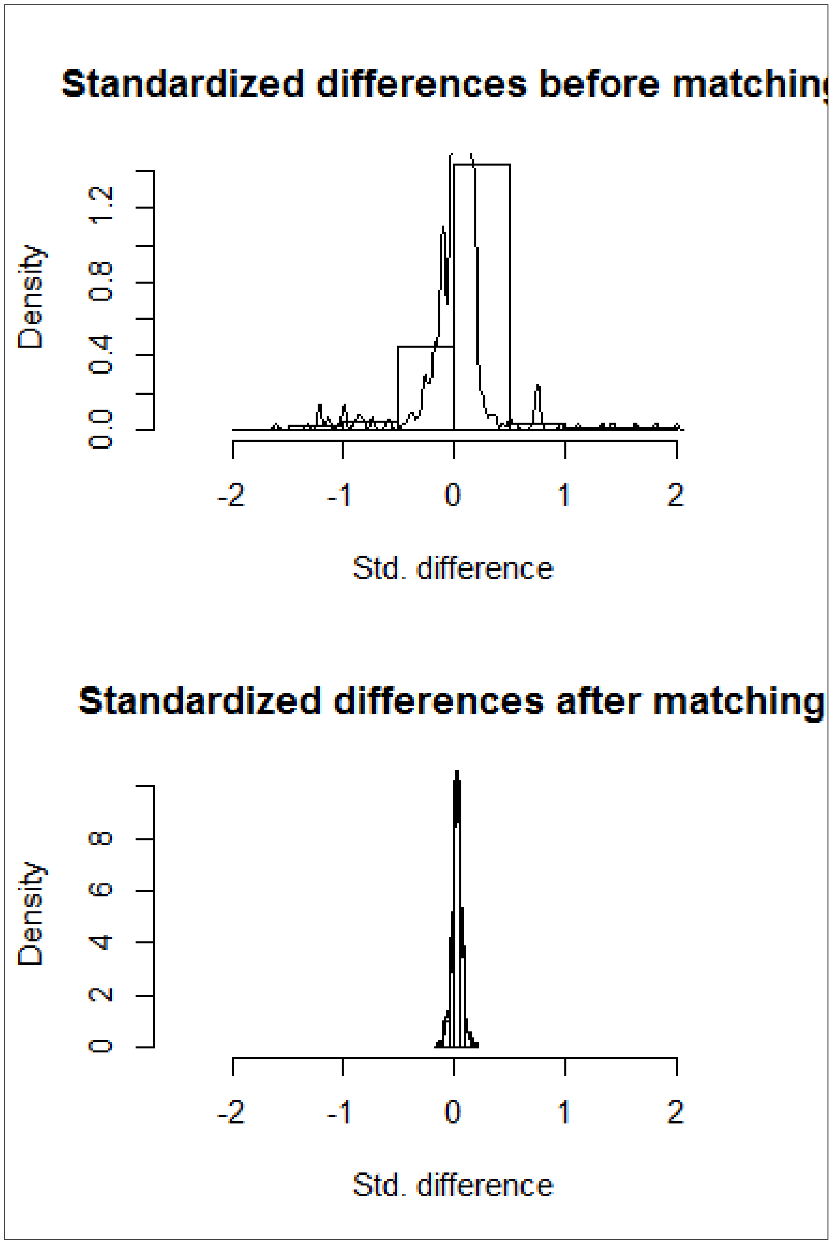

Figure 12. Dotplot of Standardized Mean Differences for Cesarean Delivery Risk Adjustment Effects Before and After Matching (n = 33,820)

Table 15

Propensity Score Matched Sample Versus the Observational Design Sample

Variable	Propensity matched sample (*n* = 33,820)		Observational design sample (*N* = 123,145)	
	OR	95% CI	*OR*	95% CI
Robson Group				
Group 5	Ref.		Ref.	
Group 1	1.27**	[1.16, 1.39]	0.02**	[0.01, 0.03]
Group 2	1.24**	[1.13, 1.36]	0.02**	[0.01, 0.03]
Group 3	1.19**	[1.08, 1.31]	0.02**	[0.02, 0.03]
Group 4	0.92	[0.82, 1.03]	0.02**	[0.01, 0.03]
Group 6	10.98**	[7.09, 15.25]	3.84**	[2.81, 5.26]
Group 7	3.53**	[2.69, 4.62]	1.64**	[1.29, 2.08]
Group 8	1.17#	[1.00, 1.36]	0.22**	[0.19, 0.25]
Group 9	3.71**	[2.70, 5.09]	0.46**	[0.34, 0.62]
Group 10	1.29**	[1.18, 1.42]	0.03**	[0.02, 0.04]
Clinical Condition				
Diabetes	1.55**	[1.43, 1.68]	1.51**	[1.40, 1.62]
Hypertension	1.23**	[1.15, 1.32]	1.95**	[1.83, 2.08]
Eclampsia	1.35**	[1.25, 1.46]	4.34**	[4.03, 4.66]
HIV	1.85**	[1.32, 2.59]	10.37**	[7.27, 14.78]
Oligohydramnios	1.44**	[1.29, 1.59]	2.06**	[1.87, 2.28]
Polyhydramnios	1.55**	[1.33, 1.80]	2.35**	[2.01, 2.76]
Fetal abnormality	1.20	[0.65, 2.21]	3.49**	[1.79, 6.82]
Antepartum	1.56**	[1.39, 1.76]	7.18**	[6.38, 8.07]
Fetal distress	1.89**	[1.74, 2.06]	16.09**	[14.81, 17.49]
Anemia	1.22**	[1.15, 1.29]	1.95**	[1.85, 2.05]
Dystocia	1.38**	[1.29, 1.46]	13.31**	[12.62, 14.03]
Fetal anomaly	1.21	[0.65, 2.24]	0.59	[0.31, 1.17]
RH	1.14	[0.99, 1.30]	1.02	[0.91, 1.15]
PROM	1.32**	[1.18, 1.47]	0.99#	[0.81, 0.99]
Genital herpes	1.37**	[1.17, 1.47]	2.27**	[1.94, 2.65]
Fetal heart	1.17**	[1.11, 1.23]	4.45**	[4.24, 4.66]
Insurance				
Public	Ref.		Ref.	
Private	0.95	[0.91, 1.00]	0.91**	[0.88, 0.95]
None	1.16*	[1.01, 1.33]	0.76**	[0.67, 0.86]
Maternal age				
30 - 35	Ref.		Ref.	
< 20	0.99	[0.92, 1.09]	0.45**	[0.41, 0.48]
20 - 24	0.90*	[0.85, 0.96]	0.62**	[0.59, 0.65]
25 - 29	0.84**	[0.80, 0.89]	0.45**	[0.73, 0.81]
> 35	1.11*	[1.03, 1.19]	1.45**	[1.36, 1.55]

Note. * = p < .01, ** = p < .001, Ref. = referent category, OR = odds ratio, CI = confidence interval.

Table 15 presents the findings of the logistic regression modeling of the matched sample (*n* = 33,820) in comparison to the observational design sample (*N* = 123,145). For clinical characteristics, the propensity matched sample suggested a moderate association with the cesarean delivery outcome in comparison to a vaginal delivery for diabetes (*OR* = 1.55, 95% CI [1.43, 1.68], $p < .001$), HIV (*OR* = 1.85, 95% CI [1.32, 2.59], $p < .001$), polyhydramnios (*OR* = 1.55; 95% CI [1.33, 1.80], $p < .001$), antepartum (*OR* = 1.56, 95% CI [1.39, 1.76], $p < .001$) and fetal distress (*OR* = 1.89, 95% CI [1.74, 2.06], $p < .001$). Notably, these findings differed from the logistic regression model in magnitude of association and confidence interval measurement. Otherwise, propensity score matching produced Robson Group effects that differed from the full model. Robson Group 6, Group 7, and Group 9 were strongly associated with the cesarean delivery procedure when compared to Group 5. For the Full model, only Group 6 demonstrated a strong association with the cesarean delivery model when compared to the referent Group 5. In the full model, Groups 1 through 4 had a strong association with a vaginal delivery when compared to Group 5. These effects were no longer noted in the propensity score matched sample.

When comparing the two models, 27 of 31 (87%) standardized model effects presented nonoverlapping confidence intervals suggesting a difference. In the comparison of p values for the propensity score matched sample, 21 of 31 effects (67%) had a p-value < .001 compared to 28 of 31 (90%) for the observational sample. Further review of comparative ROC curves, suggested a difference between the propensity score matched sample and the observational design sample (Z-score = 119.25, $p < .001$).

Stratum Specific Analyses

Research Question 7 tested whether for the propensity score matched sample, if there was a difference in cesarean delivery risk adjustment effects by stratum.

Table 16

Stratum Specific Descriptive Statistics of Matched Cesarean Treatment Cases (n = 16,910) and Vaginal Delivery Controls (n = 16,910)

		Cesarean treatment		Vaginal controls	
Strata	Propensity score range	*n*	%	*n*	%
1	0.00658 - 0.02327	106	24.5	327	75.5
2	0.02334 - 0.04095	377	42.9	501	57.1
3	0.04103 - 0.06534	465	51.5	438	48.5
4	0.06561 – 0.10027	654	44.0	833	56.0
5	0.10069 - 0.15233	1,323	49.2	1,366	50.8
6	0.15243 - 0.27759	2,736	49.9	2,747	50.1
7	0.27781 - 0.57912	5,254	48.6	5,562	51.4
8	0.57921 – 0.85946	3,164	52.8	2,824	47.2
9	0.86002 – 0.92760	509	27.7	1,326	72.3
10	0.92779 – 0.99999	2,322	70.2	986	29.8

The matched sample resulted in 16,910 cesarean delivery cases and 16,910 vaginal delivery controls. The matched sample was further stratified based on propensity score matching within stratum. The resulting 10 strata statistically differed as measured by one-way ANOVA [$F(9, 33{,}810) = 124.4$, $p = .001$)] (Table 16). Tukey's test for multiple comparisons suggested 6 of 90 strata comparisons were more likely similar because they did not reach a level of statistical significance.

Table 17

Stratum 1 Risk Adjustment Effects on the Cesarean Delivery Outcome (n = 433)

Variable	*OR*	95% CI
Robson Group		
Group 3	0.51	[0.12, 2.20]
Clinical Condition		
Diabetes	2.15	[0.68, 6.78]
Anemia	1.92	[0.87, 4.35]
RH	0.79	[0.24, 2.58]
Insurance		
Public	Ref.	
Private	1.36	[0.79, 2.32]
None	0.21	[0.03, 1.67]
Maternal age		
30 - 35	Ref.	
< 20	3.50[#]	[0.97, 12.54]
20 - 24	1.59	[0.81, 3.14]
25 - 29	0.79	[0.53, 1.79]
> 35	1.31	[0.09, 17.42]

Note. [#] = p < .05, Ref. = referent category, OR = odds ratio, CI = confidence interval.

Stratum 1 represented the lowest risk group and propensity scores ranged from 0.00658 to 0.02327 (Table 16) (Table 17). Among the 433 maternity patients, 24.5% had a cesarean procedure and 75.5% a vaginal delivery. Diabetes (*OR* = 2.15, 95% CI [0.68, 6.78]) and anemia (*OR* = 1.92, 95% CI [0.87, 4.35]) had a moderate association with the cesarean delivery procedure. Women less than 20 years of age (*OR* = 3.50, 95% CI [0.97, 12.54]) had a strong association for a cesarean delivery when compared to the referent group.

Stratum 2 included propensity scores ranging from 0.02334 to 0.04095 and represented low risk patients (Table 16) (Table 18). Among the 878 patients, approximately 42.9% had a cesarean delivery and 57.1% a vaginal delivery. RH (*OR* = 3.06, 95% CI [1.12, 8.38]) had a strong association with the cesarean delivery procedure and PROM (*OR* = 2.07, 95% CI [0.83, 5.18]), a moderate association when compared to a vaginal delivery. There were increased odds of genital herpes among women having a vaginal delivery when compared to cesarean procedures (*OR* = 0.08, 95% CI [0.01, 0.84], $p < .05$).

Stratum 3 was identified as a low risk group and propensity scores ranged from 0.04103 to 0.06534 (Table 16) (Table 19). Among 903 cases, 51.5% were cesarean procedures and 48.5% vaginal deliveries. For this subclass, Robson Groups and clinical characteristics demonstrated little to no association with the cesarean delivery outcome when compared to a vaginal delivery, whereas Robson Group 2 (*OR* = 0.25, 95% CI [0.05, 1.14]) was strongly associated with a vaginal delivery; Group 1 (*OR* = 0.43, 95% CI [0.03, 5.25]) and Group 3 (*OR* = 0.53, 95% CI [0.24, 1.18]) were moderately associated with a vaginal delivery. There was a moderate association of having a cesarean delivery among women with no insurance (*OR* = 2.09, 95% CI [0.91, 4.79]) when compared with having public insurance. Women less than 20 years of age (*OR* = 1.81, 95% CI [0.68, 4.85]), 20 to 24 y.o.a. (*OR* = 1.82, 95% CI [1.07, 3.08], $p < .05$), and 25 to 29 y.o.a (*OR* = 1.93, 95% CI [1.18, 3.17], $p < .01$) had a moderate association with the cesarean delivery outcome when compared to those 30 to 35 y.o.a.

Table 18

Stratum 2 Risk Adjustment Effects on the Cesarean Delivery Outcome (n = 878)

Variable	*OR*	95% CI
Robson Group		
Group 3	0.61	[0.29, 1.28]
Clinical Condition		
Diabetes	1.25	[0.54, 2.85]
Hypertension	0.64	[0.23, 1.80]
Oligohydramnios	0.57	[0.12, 2.79]
Fetal abnormality	0.54#	[0.13, 2.23]
RH	3.06	[1.12, 8.38]
PROM	2.07	[0.83, 5.18]
Genital Herpes	0.08#	[0.01, 0.84]
Insurance		
Public	Ref.	
Private	1.25	[0.89, 1.75]
None	0.86	[0.38, 1.92]
Maternal age		
30 – 35	Ref.	
< 20	8.21#	[0.96, 70.43]
20 - 24	0.73	[0.46, 1.16]
25 - 29	0.75	[0.52, 1.12]
> 35	0.49	[0.21, 1.12]

Note. # = p < .05, * = p < .01, ** = p < 0.001, Ref. = referent category, OR = odds ratio, CI = confidence interval.

Table 19

Stratum 3 Risk Adjustment Effects on the Cesarean Delivery Outcome (n = 903)

Variable	*OR*	95% CI
Robson Group		
Group 1	0.43	[0.03, 5.25]
Group 2	0.25	[0.05, 1.14]
Group 3	0.53	[0.24, 1.18]
Clinical Condition		
Diabetes	0.83	[0.41, 1.69]
Hypertension	0.65	[0.31, 1.32]
Eclampsia	0.51	[0.13, 1.92]
Oligohydramnios	0.85	[0.32, 2.28]
Fetal abnormality	1.15	[0.39, 3.29]
Anemia	0.64	[0.36, 1.12]
RH	0.89	[0.40, 1.94]
Fetal heart	1.17	[0.39, 3.49]
Insurance		
Public	Ref.	
Private	1.32	[0.95, 1.83]
None	2.09	[0.91, 4.79]
Maternal age		
30 - 35	Ref.	
< 20	1.81	[0.68, 4.85]
20 - 24	1.82[#]	[1.07. 3.08]
25 - 29	1.93[*]	[1.18, 3.17]
> 35	0.93	[0.60, 1.45]

Note. # = p < .05, * = p < .01, Ref. = referent category, OR = odds ratio, CI = confidence interval.

Table 20

Stratum 4 Risk Adjustment Effects on the Cesarean Delivery Outcome (n = 1,437)

Variable	*OR*	95% CI
Robson Group		
Group 1	0.97	[0.38, 2.49]
Group 2	0.98	[0.43, 2.23]
Group 3	1.23	[0.66, 2.31]
Clinical Condition		
Diabetes	1.06	[0.59, 1.89]
Hypertension	1.57	[0.91, 2.73]
Eclampsia	1.00	[0.35, 2.90]
Oligohydramnios	1.53	[0.59, 3.98]
Polyhydramnios	1.34	[0.49, 3.67]
Anemia	1.39	[0.90, 2.16]
RH	0.99	[0.54. 1.83]
PROM	0.76	[0.37, 1.58]
Genital herpes	1.35	[0.52, 3.53]
Fetal heart	2.86*	[1.39, 5.92]
Insurance		
Public	Ref.	
Private	0.78#	[0.61, 0.99]
None	1.01	[0.51, 2.21]
Maternal age		
30 - 35	Ref.	
< 20	0.56#	[0.34, 0.95]
20 - 24	0.93	[0.64, 1.34]
25 - 29	1.01	[0.70, 1.44]
> 35	0.65	[0.38, 1.09]

Note. # = p < .05, * = p < .01, Ref. = referent category, OR = odds ratio, CI = confidence interval.

Stratum 4 was classified as a low risk group, and propensity scores ranged from 0.06561 to 0.10027 (Table 16) (Table 20). Among 1,437 cases, 44.0% were cesarean and 56% vaginal deliveries. Among clinical conditions, hypertension (*OR* = 1.57, 95% CI [0.91, 2.73]) and fetal heart (*OR* = 2.86, 95% CI [1.39, 5.92], $p < .001$) had a moderate association with the cesarean procedure. Women less than 20 y.o.a (*OR* = 0.56, 95% CI [0.64, 1.34], $p < .05$) had a moderate association with the vaginal delivery procedure when compared to its referent group.

Stratum 5 was identified as a low risk group with propensity scores ranging from 0.10069 to 0.15233 (Table 16) (Table 21). Among the 2,689 cases, 49.2% were cesarean procedures and 50.8% vaginal deliveries. For this subclass, genital herpes (*OR* = 2.75, 95% CI [1.35, 5.60], $p < .01$), fetal heart (*OR* = 2.55, 95% CI [1.69, 3.87], $p < .001$), polyhydramnios (*OR* = 2.01, 95% CI [0.67, 6.06]), eclampsia (*OR* = 1.58, 95% CI [0.78, 3.19]), and PROM (*OR* = 1.65 95% CI [1.01, 2.69], $p < .05$) had a moderate association with a cesarean delivery when compared to the vaginal delivery outcome, whereas Robson Group 3 (*OR* = 1.67, 95% CI [0.89, 3.09]) and Group 4 (*OR* = 1.83, 95% CI [1.01, 3.33], $p < .05$) had a moderate association with the cesarean delivery procedure. Having no insurance (*OR* = 1.92, 95% CI [1.15, 3.19], $p < .01$) and women greater than 35 years of age (*OR* = 1.58, 95% CI [1.04, 2.39], $p < .05$) had a moderate association with the cesarean delivery procedure. Fetal distress was strongly associated with a vaginal delivery (*OR* = 0.22, 95% CI [0.02, 1.98]).

Table 21

Stratum 5 Risk Adjustment Effects on the Cesarean Delivery Outcome (n = 2,689)

Variable	*OR*	95% CI
Robson Group		
Group 1	1.38	[0.89, 2.15]
Group 2	1.39	[0.83, 2.34]
Group 3	1.67	[0.89, 3.09]
Group 4	1.83#	[1.01, 3.33]
Clinical Condition		
Diabetes	1.47	[0.99, 2.17]
Eclampsia	1.58	[0.78, 3.19]
Oligohydramnios	0.88	[0.56, 1.40]
Polyhydramnios	2.01	[0.67, 6.06]
Fetal abnormality	1.09	[0.52, 2.29]
Antepartum	0.67	[0.21, 2.14]
Anemia	0.83	[0.61, 1.13]
Fetal distress	0.22	[0.02, 1.98]
RH	1.39	[1.35, 5.60]
PROM	1.65#	[1.01, 2.69]
Genital herpes	2.75*	[1.35, 5.60]
Fetal heart	2.55**	[1.69, 3.87]
Insurance		
Public	Ref.	
Private	0.84	[0.70, 1.01]
None	1.92*	[1.15, 3.19]
Maternal age		
30 - 35	Ref.	
< 20	0.71	[0.47, 1.06]
20 - 24	0.72*	[0.56, 0.91]
25 - 29	0.76*	[0.62, 0.94]
> 35	1.58#	[1.04, 2.39]

Note. # = p < .05, * = p < .01, ** = p < .001, Ref. = referent category, OR = odds ratio, CI = confidence interval.

With propensity scores ranging from 0.15243 to 0.27759, Stratum 6 was defined as a low risk group (Table 16) (Table 22). This stratum included 5,483 cases with 49.9% cesarean and 50.1% vaginal deliveries. Robson Group 3 (*OR* = 2.28, 95% CI [1.34, 3.88], $p < .01$) had a moderate association with the cesarean delivery, and Robson 4 (*OR* = 3.87, 95% CI [1.75, 8.47], $p < .001$) indicated a strong association.

Stratum 7 was the largest in size with 10,816 observations, and propensity scores ranged from 0.27781 to 0.57912 (Table 16) (Table 23). Fetal distress (*OR* = 3.37, 95% CI [2.39. 4.78], $p < .001$) had a strong association with the cesarean delivery procedure. Other clinical conditions with a moderate association with the cesarean deliver procedure included antepartum (*OR* = 2.47, 95% CI [1.85, 3.58], $p < .001$), HIV (*OR* = 2.34, 95% CI [1.30, 4.19], $p < .01$), eclampsia (*OR* = 1.60, 95% CI [1.33, 1.94], $p < .001$), and dystocia (*OR* = 2.81, 95% CI [2.11, 3.75], $p < .001$). Robson Group 3 (*OR* = 0.21, 95% CI [0.12, 0.38], $p < .001$) and Group 4 (*OR* = 0.16, 95% CI [0.08, 0.32], $p < .001$) were strongly associated with a vaginal procedure, and Robson Group 1 (*OR* = 0.47, 95% CI [0.28, 0.79], $p < .01$) and Group 2 (*OR* = 0.40, 95% CI [0.23, 0.69], $p < .001$) were moderately associated with this outcome.

Table 22

Stratum 6 Risk Adjustment Effects on the Cesarean Delivery Outcome (n = 5,483)

Variable	*OR*	95% CI
Robson Group		
Group 1	1.19	[0.96, 1.48]
Group 2	1.31	[0.95, 1.82]
Group 3	2.28*	[1.34, 3.88]
Group 4	3.87**	[1.75, 8.47]
Clinical Condition		
Diabetes	0.82	[0.64, 1.03]
Eclampsia	0.44**	[0.28, 0.68]
Oligohydramnios	0.59*	[0.43, 0.83]
Polyhydramnios	0.78	[0.45, 1.37]
Fetal abnormality	0.76	[0.24, 5.00]
Antepartum	0.73	[0.39, 1.39]
Anemia	0.66**	[0.54, 0.81]
Fetal distress	0.92**	[0.40, 0.21]
RH	0.75	[0.52, 1.06]
PROM	1.10	[0.85, 1.42]
Genital herpes	1.04	[0.68, 1.58]
Fetal heart	0.78	[0.54, 1.15]
Insurance		
Public	Ref.	
Private	0.88#	[0.79, 0.99]
None	0.79	[0.54, 1.18]
Maternal age		
30 - 35	Ref.	
< 20	1.09	[0.82, 1.47]
20 - 24	0.89	[0.79, 1.07]
25 - 29	1.03	[0.87, 1.27]
> 35	0.98	[0.79, 1.24]

Note. # = p < .05, * = p < .01, ** = p < .001, Ref. = referent category, OR = odds ratio, CI = confidence interval.

Table 23

Stratum 7 Risk Adjustment Effects on the Cesarean Delivery Outcome (n = 10,816)

Variable	*OR*	95% CI
Robson Group		
Group 1	0.47*	[0.28, 0.79]
Group 2	0.40**	[0.23, 0.69]
Group 3	0.21**	[0.12, 0.38]
Group 4	0.16**	[0.08, 0.32]
Clinical Condition		
Diabetes	1.10	[0.96, 1.26]
Hypertension	1.24*	[1.08, 1.43]
Eclampsia	1.60**	[1.33, 1.94]
HIV	2.34*	[1.30, 4.19]
Oligohydramnios	1.22*	[1.01, 1.48]
Polyhydramnios	1.21	[0.90, 1.62]
Fetal abnormality	1.31	[0.54, 0.81]
Antepartum	2.47**	[1.85, 3.58]
Anemia	1.15*	[1.02, 1.28]
Dystocia	2.81**	[2.11, 3.75]
Fetal anomaly	1.04	[0.38, 2.85]
Fetal distress	3.37**	[2.39, 4.78]
RH	0.91	[0.71, 1.15]
PROM	1.08	[0.88, 1.31]
Genital herpes	0.82	[0.60, 1.11]
Fetal heart	1.32	[1.10, 1.58]
Insurance		
Public	Ref.	
Private	0.96	[0.88, 1.04]
None	0.89	[0.68, 1.16]
Maternal age		
30 - 35	Ref.	
< 20	0.84#	[0.71, 0.99]
20 - 24	0.98	[0.87, 1.09]
25 - 29	1.03	[0.87, 1.27]
> 35	1.29**	[1.13, 1.49]

Note. # = p < .05, * = p < .01, ** = p < .001, Ref. = referent category, OR = odds ratio, CI = confidence interval.

Stratum 8 was identified as a moderate to high risk group with propensity scores ranging from 0.57921 to 0.85946 (Table 16) (Table 24). This class included 5,988 cases with 52.8% being cesarean procedures and 47.2% vaginal deliveries. Diabetes (*OR* = 1.50, 95% CI [1.25, 1.80], $p < .001$), hypertension (*OR* = 2.01, 95% CI [1.68, 2.40], $p < .001$), oligohydramnios (*OR* = 2.12, 95% CI [1.63, 2.72], $p < .001$), fetal abnormality (*OR* = 2.15, 95% CI [1.57, 2.93], $p <.001$), anemia (*OR* = 1.78, 95% CI [1.54, 2.05], $p < .001$), RH (*OR* = 1.79, 95% CI [1.31, 2.46], $p < 0.001$), and PROM (*OR* = 1.82, 95% CI [1.44, 2.29], $p < .001$) had a moderate association with the cesarean delivery outcome when compared to a vaginal delivery.

Stratum 9 was classified as a high risk group, and propensity scores ranged from 0.86002 to 0.92760 (Table 16) (Table 25). Among the 1,835 observations, 27.7% were cesarean and 72.3% vaginal deliveries. Diabetes (*OR* = 3.27, 95% CI [2.14, 4.98], $p < .001$), eclampsia (*OR* = 3.79, 95% CI [2.41, 5.98], $p < .001$), and polyhydramnios (*OR* = 3.39, 95% CI [1.76, 6.94], $p < .001$) had a strong association with the cesarean delivery procedure. Clinical conditions with a moderate association with the cesarean procedure included hypertension (*OR* = 2.69, 95% CI [1.75, 4.14], $p < .001$), anemia (*OR* = 2.39, 95% CI [1.75, 3.26], $p < .001$), fetal anomaly (*OR* = 2.98, 95% CI [1.45, 6.13], $p < .05$), PROM (*OR* = 2.76, CI [1.51, 5.03], $p < .001$), and genital herpes (*OR* = 2.64, 95% CI [0.97, 7.18], $p < .05$). Having no insurance was moderately associated with a cesarean delivery when compared to public insured maternity patients (*OR* = 2.93, 95% CI [1.61, 5.31], $p < .001$). Maternity patients less than 20 years of age were strongly associated with the cesarean delivery procedure when compared to its referent group (*OR* = 5.61, 95% CI [3.29, 9.59], $p < .001$). Robson Group 3 (*OR* = 0.26, 95% CI [0.09, 0.73], $p < .05$) and Group 5 (*OR* = 0.21, 95% CI [0.11, 0.39], $p < .001$) were strongly associated with the vaginal delivery. For the latter group, this may represent women who had a vaginal delivery after a previous cesarean delivery.

Stratum 10 represented the highest risk group, with propensity scores ranging from 0.92779 to 0.99999 (Table 16) (Table 26). Among the 3,308 events, 70.2% were cesarean procedures and 29.8% vaginal deliveries. Robson Group 6 (*OR* = 10.49, 95% CI [7.07, 15.55], $p < .001$), Group 7 (*OR* = 3.43, 95% CI [2.38, 4.93], $p < .001$) and Group 9 (*OR* = 3.21, 95% CI [1.73, 5.96], $p < .05$) had a strong association with the cesarean delivery outcome when compared to Group 5. Whereas Group 1 (*OR* = 0.18, 95% CI [0.12, 0.29], $p < .001$), Group 2 (*OR* = 0.06, 95% CI [0.03, 0.09]), Group 3 (*OR* = .02; 95% CI [0.01, 0.03], $p < .001$), Group 4 (*OR* = 0.02, 95% CI [0.01, 0.03], $p < .05$), Group 8 (*OR* = 0.29, 95% CI [0.18, 0.45], $p < .001$) and Group 10 (*OR* = 0.06, 95% CI [0.04, 0.11], $p < .001$) demonstrated a strong association with the vaginal delivery outcome when compared to the referent Group 5.

The majority of clinical conditions were strongly associated with the cesarean delivery procedure, with point estimates exceeding an odds ratio of 3.00. Hypertension (*OR* = 2.05, 95% CI [1.55, 2.75]), anemia (*OR* = 2.31, 95% CI [1.83, 2.91], $p < .001$) and PROM (*OR* = 1.54, 95% CI [0.96, 2.48]) were moderately associated with the cesarean procedure. Having private (*OR* = 1.68, 95% CI [1.37, 2.05], $p < .001$) or no insurance (*OR* = 1.53, 95% CI [0.87, 2.69]) were moderately associated with the cesarean delivery procedure when compared to the public insurance referent group.

Table 24

Stratum 8 Risk Adjustment Effects on the Cesarean Delivery Outcome (n = 5,988)

Variable	*OR*	95% CI
Robson Group		
Group 1	1.08	[0.90, 1.30]
Group 2	1.09	[0.88, 1.35]
Group 3	1.02	[0.72, 1.43]
Group 4	0.43**	[0.26, 0.71]
Group 5	0.68	[0.39, 1.18]
Group 8	0.68#	[0.48, 0.96]
Group 9	2.39*	[1.38, 4.18]
Clinical Condition		
Diabetes	1.50**	[1.25, 1.80]
Hypertension	2.01**	[1.68, 2.40]
Eclampsia	1.15	[0.89, 1.48]
HIV	1.29	[0.66, 2.53]
Oligohydramnios	2.12**	[1.63, 2.72]
Polyhydramnios	1.33	[0.96, 1.85]
Fetal abnormality	2.15**	[1.57, 2.93]
Antepartum	1.13	[0.79, 1.84]
Fetal distress	1.51	[1.01, 2.26]
Anemia	1.78**	[1.54, 2.05]
Dystocia	0.97	[0.68, 1.40]
RH	1.79**	[1.31, 2.46]
PROM	1.82**	[1.44, 2.29]
Genital herpes	1.06	[0.72, 1.54]
Fetal heart	0.81	[0.65, 1.01]
Insurance		
Public	Ref.	
Private	0.71**	[0.63, 0.79]
None	1.75**	[1.25, 2.45]
Maternal age		
30 - 35	Ref.	
< 20	1.21	[0.95, 1.54]
20 - 24	0.89	[0.76, 1.05]
25 - 29	0.65**	[0.56, 0.76]
> 35	0.90	[0.75, 1.09]

Note. # = $p < .05$, * = $p < .01$, ** = $p < .001$, Ref. = referent category, *OR* = odds ratio, CI = confidence interval.

Table 25

Stratum 9 Risk Adjustment Effects on the Cesarean Delivery Outcome (n = 1,835)

Variable	*OR*	95% CI
Robson Group		
Group 1	1.06	[0.56, 1.92]
Group 2	0.44$^{\#}$	[0.22, 0.89]
Group 3	0.26$^{\#}$	[0.09, 0.73]
Group 5	0.21**	[0.11, 0.39]
Group 8	0.42*	[0.21, 0.84]
Group 10	0.43*	[0.22, 0.84]
Clinical Condition		
Diabetes	3.27**	[2.14, 4.98]
Hypertension	2.69**	[1.75, 4.14]
Eclampsia	3.79**	[2.41, 5.98]
Polyhydramnios	3.39**	[1.76, 6.94]
Antepartum	1.55	[0.83, 2.88]
Fetal distress	1.51	[0.91, 2.52]
Anemia	2.39**	[1.75, 3.26]
Dystocia	0.56*	[0.35, 0.89]
Fetal anomaly	2.98$^{\#}$	[1.45, 6.13]
PROM	2.76**	[1.51, 5.03]
Genital herpes	2.64$^{\#}$	[0.97, 7.18]
Fetal heart	1.07	[0.71, 1.59]
Insurance		
Public	Ref.	
Private	1.06	[0.79, 1.42]
None	2.93**	[1.62, 5.31]
Maternal Age		
30 - 35	Ref.	
< 20	5.61**	[3.29, 9.59]
20 - 24	1.52$^{\#}$	[1.04, 2.20]
25 - 29	0.77	[0.54, 1.09]
> 35	2.99**	[1.91, 4.68]

Note. $^{\#}$ = *p* < .05, * = *p* < .01, ** = *p* < .001, Ref = referent category, *OR* = odds ratio, CI = confidence interval.

Table 26

Stratum 10 Risk Adjustment Effects on the Cesarean Delivery Outcome (n = 3,308)

Variable	*OR*	95% CI
Robson Group		
Group 5	Ref.	
Group 1	0.18**	[0.12, 0.29]
Group 2	0.06	[0.03, 0.09]
Group 3	0.02**	[0.01, 0.03]
Group4	0.02#	[0.02, 0.03]
Group 6	10.49**	[7.07, 15.55]
Group 7	3.43**	[2.38, 4.93]
Group 8	0.29**	[0.18, 0.45]
Group 9	3.21#	[1.73, 5.96]
Group 10	0.06**	[0.04, 0.11]
Clinical Condition		
Diabetes	9.16**	[7.00, 11.99]
Hypertension	2.05	[1.55, 2.75]
Eclampsia	6.70**	[4.70, 9.50]
Oligohydramnios	5.55**	[3.62, 8.51]
Polyhydramnios	5.36**	[3.26, 8.82]
Antepartum	6.12**	[3.97, 9.45]
Fetal distress	12.92**	[9.24, 18.07]
Anemia	2.31**	[1.83, 2.91]
Dystocia	6.25**	[4.74, 8.23]
Fetal anomaly	3.79*	[2.38, 6.03]
RH	1.42	[0.77, 2.62]
PROM	1.54	[0.96, 2.48]
Genital herpes	5.98	[3.64, 9.85]
Fetal heart	1.26**	[0.99, 1.58]
Insurance		
Public	Ref.	
Private	1.68**	[1.37, 2.05]
None	1.53	[0.87, 2.69]
Maternal Age		
30 - 35	Ref.	
< 20	1.11**	[0.65, 1.90]
20 - 24	0.71	[0.52, 0.97]
25 - 29	0.54	[0.42, 0.71]
> 35	0.86**	[0.68, 1.09]

Note. # = p < .05, * = p < .01, ** = p < .001, Ref. = referent category, OR = odds ratio, CI = confidence interval.

ROC findings suggested a difference in the comparison of stratum specific effects (Table 27). For example, Strata 7, 8, 9, and 10 comparisons significantly differed from other strata at the $p < .001$ level. For the 90 ROC comparisons generated, and excluding values of 1.0, only eight comparisons (8.8%) were not statistically significant.

Table 27

Stratum Specific Comparison of Receiver Operating Characteristic Z-Scores

Strata	1	2	3	4	5	6	7	8	9	10
1	1.0	0.92	11.92**	6.49**	10.02**	2.06**	6.86**	16.97**	76.35**	65.37**
2	0.92	1.0	12.84**	7.41**	10.94**	1.14	5.94**	17.90**	77.34**	66.35*
3	11.92**	12.84**	1.0	5.40**	1.89	14.00**	18.82**	5.03**	63.74**	52.91**
4	6.49**	7.41**	5.40**	1.0	1.87	8.56**	18.82**	10.45**	69.41**	58.53**
5	10.02**	10.94**	1.89	1.87	1.0	12.09**	16.91**	69.20**	65.71**	54.86**
6	2.06**	1.14	14.00**	8.56**	12.09**	1.0	4.79**	19.05**	78.59**	67.58**
7	6.86**	5.94**	18.82**	13.37**	16.91**	4.79**	1.0	23.90**	83.87**	72.77**
8	16.97**	17.90**	5.03**	10.45**	69.20**	19.05**	23.90**	1.0	58.55**	47.75**
9	76.35**	77.34**	63.74**	69.41**	65.71**	78.59**	83.87**	58.55**	1.0	10.94**
10	65.37**	66.35**	52.91**	58.53**	54.86**	67.58**	72.77**	47.75**	10.94**	1.0

Note: ** = $p < .001$.

Summary

For Research Question 1, the null hypothesis was rejected and the alternative accepted because the mean cesarean delivery rate was statistically different among Robson 10 Groups.

Four hypothesis questions were proposed for cesarean delivery hospital profiling. For Research Question 2, the null hypothesis was accepted and the alternative hypothesis rejected because there was no statistical difference in standardized risk adjustment effects when comparing the cesarean delivery logistic regression versus HGLM models. For Research Question 3, the null hypothesis was rejected and the alternative hypothesis accepted because there was a statistical difference in standardized risk adjustment effects when comparing the cesarean delivery logistic regression propensity score risk adjustment versus HGLM propensity score risk adjustment models. For Research Question 4, the null hypothesis was rejected and the alternative hypothesis was accepted because cesarean delivery hospital ranks statistically differed when comparing logistic regression versus HGLM models. Whereas, for Research Question 5, the null hypothesis was rejected and the alternative hypothesis accepted because there was a statistical difference in cesarean delivery hospital ranks when comparing logistic regression propensity score risk adjustment with HGLM propensity score risk adjustment models.

Two hypotheses were proposed for propensity score matching with stratification. For Research Question 6, the null hypothesis was rejected and the alternative hypothesis accepted because there was a statistical difference in standardized risk adjustment effects when comparing the cesarean delivery propensity score matched sample versus the observational design sample. For Research Question 7, the null hypothesis was rejected and the alternative hypothesis accepted because there was a statistical difference in cesarean delivery risk adjustment effects by stratum. The next chapter advances the discussion, providing conclusions and recommendations based on findings from the research.

Chapter 5: Discussion, Conclusions, and Recommendations

Introduction

The World Health Organization advocated for the appropriate use of the cesarean deliver procedure and population-based rates ranging from 10 to 15% (Ye, Betran, Vela, Souza, & Zhang, 2014). Despite international calls for the reduction and appropriate use of the cesarean delivery procedure, Georgia had a 2012 statewide average of approximately 35.1% live births being by cesarean. Significant variation existed among the 86 hospitals examined with the cesarean delivery unadjusted rate, ranging from a low of 12.1% to a high of 48.5%. Over 50 hospitals had a cesarean delivery rate greater than 30% and a proportion of these procedures may represent medically unjustifiable procedures. This dissertation applied comparative risk adjustment models and statistical techniques for the evaluation of the cesarean delivery outcome and hospital ranks using a population-based linked dataset. Discriminating variation using differing techniques provided insights regarding potential biases in existing hospital ranks, introduced new ways of examining risk patterns according to propensity scores, and emphasized the opportunity for reducing variation in cesarean delivery rates as guided by statewide methodologies.

Interpretation of Findings

Robson 10 Group Classification Indicators

1. RQ 1: Is there a difference in the mean cesarean delivery rate among the Robson 10 Groups?

No U.S. study used population-based linked data for the extraction and use of Robson indicators in cesarean delivery risk adjustment and hospital profiling methods. Robson's 10 Group Classification has gained increased attention and endorsement at the global level as a monitoring and quality improvement tool for use at differing levels of the health care system (WHO, 2015). This research demonstrated Robson Groups were readily extracted from the 2012 linked data and represented discrete group differences in population-based rates. These findings were attributed to the completeness of the live birth file and availability of other clinical variables ascertained from the hospital discharge summary file. Other population-based findings suggested high quality live birth data would exist for relevant variables in creating the Robson Classification System, including previous cesarean delivery, cesarean method of delivery, and cephalic presentation (Martin et al., 2013). Study findings from this research demonstrated previous cesarean delivery, breech, abnormal lie positions, and cephalic presentation were not listed in the live birth file, requiring the use of hospital discharge summary data elements instead. Similar to Kahn et al. (2009), the majority of clinical characteristics was ascertained from the hospital discharge summary file and may represent higher quality data in comparison to the live birth file.

Significant variation in Robson Groups existed across hospitals, further demonstrating the need for standard clinical practice aimed at reducing variation where appropriate, and recognizing volume and geographical place may be driving a proportion of rates versus clinical and SES characteristics alone (Kozhimannil et al., 2013). Absent from this research was an analysis of the association of maternal and infant outcomes versus cesarean section rates in each of the 10 groups for an understanding of optimal rates in comparison to outcomes (Betran et al., 2014). The accuracy of this information depended on further linkages to 2012 Georgia fetal death, infant death, and maternal death records, and these were unavailable at the time of research.

Hospital Profiling

2. RQ 2: Is there a difference in cesarean delivery risk adjustment effects when comparing logistic regression versus hierarchical generalized linear models?
3. RQ 3: Is there a difference in cesarean delivery risk adjustment effects when comparing logistic regression propensity score versus hierarchical generalized linear propensity score models?
4. RQ 4: Is there a difference in cesarean delivery hospital ranks when comparing logistic regression versus hierarchical generalized linear models?
5. RQ 5: Is there a difference in cesarean delivery hospital ranks when comparing logistic regression propensity score risk adjustment versus hierarchical generalized linear model propensity score risk adjustment?

No U.S. study used comparative statistical methods in the examination of cesarean delivery risk adjustment, propensity score risk adjustment, and hospital ranks. Findings from this research suggests model effects and hospital ranks varied by risk adjustment variable selection, whether propensity scores were included in modelling or not, and by statistical technique. These noted differences could influence markedly the interpretation of association of the cesarean delivery outcome and appraisal of evidence when in support of decision making or policy. For example, a clinical characteristic noted as being strongly associated with the outcome in one model could have a weak association in another. This is an important distinction to describe because reproductive health studies often relied on logistic regression methodologies (Bailit & Garrett, 2003), with limited application of provider profiling multilevel models (Leung et al., 1998) or propensity score approaches (Gilbert et al., 2012).

Irrespective of statistical technique or risk adjustment method, only fetal distress was strongly associated with the cesarean delivery outcome in four of the six models (Table 6). Fetal distress was moderately associated with the cesarean delivery procedure in five of the six models. Anemia and hypertension were moderately associated with the cesarean delivery outcome in at least three of the six models. Fetal anomaly, RH, PROM, diabetes, hypertension, oligohydramnios, and polyhydramnios had a weak association with the cesarean delivery in at least three of six models reviewed. These findings were not wholly congruent with the risk adjustment findings by Bailet (2003) and may reflect a difference in statistical technique and risk adjustment models used. Bailet (2003) created clinical composite variables when modeling cesarean delivery effects, defined as severe, moderate, and mild. Whereas, this research modelled each of the clinical characteristics comprising the three composite risk categories, and were entered into statistical models as independent variables. Notwithstanding, this research suggested there were few clinical conditions having a strong or moderate association with the cesarean delivery procedure implying a degree of inappropriate utilization at the population level.

Few studies modelled Robson 10 Groups as a risk adjustor in the examination of the cesarean delivery outcome or hospital ranks. Like other researchers, model effects differed according to risk adjustor type, statistical technique, and whether the Robson Groups were modelled alone or other maternal and demographic variables were included (Maso et al., 2013). Unlike other researchers which selected Robson Group 1 as the referent group, this research selected Robson Group 5 instead. The rational for using Group 5 is that it represented a high risk for cesarean delivery for the state of Georgia. Modeling approaches that first applied Group 1 as the referent, however, the plausibility of results were highly questionable and may reflect the structure of the data or statistical error.

In the review of hospital ranks, the risk for a cesarean delivery significantly varied among facilities according to risk adjustment model and statistical technique (Maso et al., 2013; Bailit et al., 1999; Leung et al., 1998; Coonrod et al., 2008). Findings from this research suggest hospital ranks generated from HGLM models statistically differed from the logistic regression models. Few provider profiling studies have applied propensity scores as a risk adjustor to multilevel models in the study of health outcomes. Findings from this research demonstrated hospital ranks did not significantly differ when comparing the HGLM Full Model with the propensity score or with ones excluding it as a risk adjustor.

Previous researchers having developed provider profiling methods applied multinomial regression techniques and modeled up to 20 hospitals outcomes (Huang, Frangakis, Dominici, et al., 2005; Shahian & Normand, 2008). Due to the inability to model 86 hospital outcomes for the statewide approach, and with limitations to existing statistical packages available at the time of research, binary outcome models were applied instead. Hospital profiling studies comparing multilevel modeling of a binary outcome versus multinomial regression outcomes models were not identified as guidance to this research.

Propensity Score Matching

6. RQ 6: Is there a difference in cesarean delivery risk adjustment effects when comparing the propensity score matched sample versus the observational design sample?
7. RQ 7: For the propensity score matched sample, is there a difference in cesarean delivery risk adjustment effects by stratum?

No U.S. study applied cesarean delivery propensity score matching with stratification to the observational design. The matched sample of cesarean delivery cases and vaginal controls allowed for the examination of observed effects of a nonrandomized design, given the difficulties of randomizing women to a cesarean or vaginal delivery procedure. The matched sample's propensity score was based on data reduction techniques for the creation of a unique patient predicted probability and use in one-to-one matching to achieve the balanced design. Stratifying the propensity score matched sample into subclasses allowed for an examination of the distribution of cesarean risk across 10 strata (Saha et al., 2013; Kurth et al., 2005). Given the critical need to examine more closely variations of the cesarean delivery for a matched design, an assumption was made that the higher the propensity scores by stratum, the greater the risk for cesarean delivery.

Typically, researchers applied propensity score stratification to the total observed sample. Known approaches equally divide the sample into quintiles, with the assumption that 90% of the biased is eliminated (Rosenbaum & Ruben, 1983). For this research, the propensity score matching was first applied to the total observed sample (*N* = 123,145). From the one-to-one case (treatment) and control matches (*n* = 33,820), 10 groups were created using stratification procedures. Within each group, a case (treatment) and a control were matched based on similarity in the propensity score.

When comparing the propensity score matched sample (*n* = 33,820) to the logistic regression observational sample (*N* = 123,145), noted differences existed. Among clinical characteristics, the logistic regression full model presented point estimates and confidence intervals that may be inflated for oligohydramnios, fetal distress, dystocia, fetal heart, and HIV. These findings may be due to a small number of events distributed among the 86 hospitals. Among the propensity score matched sample, all clinical characteristics were either of moderate or weak association with the cesarean delivery outcome. For the logistic regression full model (*N* = 123,145), 10 clinical characteristics have a strong association with the cesarean delivery outcome, three a moderate association, and two a weak association. Noted differences also existed in the comparison of Robson Groups. For the propensity score matched sample, Group 2 and Groups 6 -10, the log odds were markedly different from the logistic regression full model. Among clinical characteristics, the log odds ratio of the PS matched model (*n* = 33,820) was comparable to the PS HGLM full model (*N* =123,415).

By creating a propensity score matched sample with stratification, this research demonstrated an increase likelihood of inappropriate utilization of the cesarean delivery especially among low risk groups having comprised over 80% of patients. Given the global escalation of cesarean delivery rates and questionable utilization practices, this research presented a new strategy of segmenting risk according to propensity classes based on the match cases and controls.

The findings from this research demonstrated the importance of conducting comparative analysis of the cesarean delivery outcome by risk adjustment model and statistical technique. The propensity score matching with stratification suggested the highest risk group of women, representing 10% of the study sample, more likely had appropriate utilization of the cesarean delivery. Even then, a proportion of women delivered by vaginal delivery, and cesarean deliveries were averted. Among hospital profiling methods, the hierarchical generalized linear models were more likely the valid approach because these models account for the nested features of the data in comparison to conditional logistic regression. Increasingly, hierarchical linear models were endorsed when the aim is to describe variation in practice patterns at the population level or when causal effects may be measured at differing levels and in accordance with the nested features of the data (Houchens, Chu, & Steiner, 2007).

Study Limitations

A number of study limitations were noted as related to administrative data sources, risk adjustment methods, and statistical technique. For certain variables, the original data sources were of medium to poor quality, and this aligned with findings from other studies (Martin et al., 2013). As noted, the 2012 Georgia live birth file was incomplete and did not include a number of key maternal and infant clinical characteristics, which limited the opportunity for external validation with the 2012 hospital discharge summary file. In designing propensity score models, it is important to include as many possible confounders in alignment with clinical guidelines, research evidence, and based on the quality of existing data sources. The propensity score models used for this research may be biased due to unaccounted variables that were not included in the model, either because of poor quality data or due to a lack of measurement in the administrative data sources (Austin, Mamdani, Stukel, Anderson, & Tu, 2005; Gregory, Korst, Gornbein, & Platt, 2002). For example, the obesity variable was eliminated from the analysis because it was poorly represented and not at an expected rate when ascertained from the hospital discharge summary data. Yet, body mass index is routinely collected through the live birth file and may be associated with the cesarean delivery procedure (DeClercq, MacDorman, Osterman, Belanoff, & Iverson, 2015). Other clinical conditions associated with the cesarean delivery outcome may not have been at the expected rate or quality and for similar distribution across facilities. This scenario would affect in particular the hierarchical linear models due to sparse numbers and the nesting features of the data (Bell, Ferron, & Kromrey, 2008). Other relevant clinical characteristics were unmeasured and may be associated with the cesarean delivery and better ascertained via clinical records (Zhang et al., 2010). Also absent from this research were deliveries that occurred in federal hospitals, or out of hospital births, because they were not a part of the 2012 Georgia Hospital Discharge Summary file, and may represent less than 5% of the sample.

Propensity score matching with stratification was not well studied through simulation or comparative models. Stratum specific analyses were common in epidemiologic studies for the examination of confounding, effect modification, and additive assumptions. Logistic and hierarchical generalized linear methods used conditional methods on the stratum specific effects, and it was not clear regarding the magnitude of bias resulting from propensity score matching with stratification. It was also not clear whether other statistical techniques of modeling on the stratum specific marginal effects would produce more accurate point estimates and confidence intervals. Even with comparative models, the researcher has to determine adequate statistical, risk adjustment, and propensity score techniques. This research demonstrated comparative approaches may be used with some similarities in effects and ranks, yet there were differences based on model development and statistical technique used. Studies having compared hospital profiling or propensity score models for a binary outcome, typically examined point estimates, confidence intervals, correlation coefficients or p values for an understanding of whether there was a significant difference in standardized model effects (Alexandrescu, 2011; Austin, 2003; Kurth, 2005). However, these methods did not statistically quantify the magnitude of the difference in the comparison of individual or model effects.

A study of physician profiling of multinomial outcomes advanced comparative model approaches for risk adjustment and statistical technique (Huang, Frangakis, Dominici, 2005). Yet, few studies were identified having applied these methods or validated using simulation approaches. This research was also challenged by statistical packages used and could not achieve the methods advanced by Huang, Frangakis, Dominici (2005) given their modeling a multinomial outcome of only 20 providers versus this research required a similar approach for 86 hospitals. Greater application of statistical simulation methodologies may assist in determining the validity of provider profiling research and true values. For the reproductive health area, there was limited simulation methods identified related to hospital profiling or propensity score matching, with the majority of core work from the cardiovascular area. Although methodological approaches were generated by other research areas, it is questionable if those methods were generalizable among study designs.

Recommendations for Future Research

Study limitations suggested areas of needed research in the study of statewide studies of provider profiling and propensity score methods and their application to the cesarean delivery outcome.

The majority of U.S. studies that examined population-based provider profiling methodologies for the cesarean delivery outcome are dated. Since then, methodological approaches have advanced through improved data linking strategies, multilevel modeling, and propensity score approaches. Simulation statistical models examining propensity score matching with stratification conditioned on the cesarean delivery outcome should inform applied research. Comparative models examining provider profiling methods using risk adjustment methods in the study of variation as compared to multinomial regression outcomes would assist in advancing model specifications. Conducting longitudinal hospital profiling methodologies would aid in understanding variation in effects having attributed to an increase in cesarean delivery rates over time.

Timely quality improvement research on the structural effects (Donabedian, 2002) driving the cesarean intra- and interfacility cesarean delivery rates for an understanding of practice variation and decision making, are needed. Assurances for the population-based collection of data from birthing centers and midwifery practices are required in order to examine their effects on the cesarean delivery outcome in comparison to predominant physician models. Robson 10 Group Classification System's introduction into national and statewide quality improvement programs would assist in the systematic collection of population-based indicators and for their surveillance and research.

Implications for Social Change

Current cesarean delivery rates reflect the establishment of new health care norms which, over time, have increased the likelihood of inappropriate utilization of this surgical procedure. In an effort to assure appropriate use, the WHO endorses the Robson Group Classification for the routine monitoring and comparison of the cesarean delivery within facilities, between facilities, and at the population level. Although there is a call for the reduction of the cesarean delivery rate at the global and national levels, greater attention is needed at the state and facility level through public health, private organizations, and communities. Recognizing the inappropriate use of the cesarean delivery procedure as a public health problem may place a greater urgency on reducing rates, as well as instituting policies and guidelines to hold health care reform to the necessary accountability. Mandating the statewide use of Robson's 10 Group Classification indicators could advance significantly cesarean delivery quality improvement initiatives, their standard application for comparisons, and timely response.

The statewide study of cesarean delivery variation could produce new approaches for examining hospital ranks according to risk adjustment methods and statistical techniques. Quality improvement programs for examining cesarean delivery adjusted rates, according to ranks, are required to advance active and timely review of information for response. As advanced as the U.S. health care system is, an increase in inappropriate rates is a new medical norm reflective of the skill and organization of practicing clinicians and of the health systems in which they treat their patients.

In 1910, it was recognized that appropriate cesarean delivery utilization was dependent upon skilled birth attendants (Spalding, 1910). It is a contradiction to assume an increase in inappropriate utilization was attributed to unskilled birth attendants when it is recognized that the U.S. health care system, and the education systems which support it, are highly regarded. Rather, the unskilled nature of clinical practice and health system patterns may be relevant as an area of needed quality review given change in organizational practice and societal norms over time. More recent initiatives demonstrated practice variation in the cesarean delivery outcome can be reduced through facility based quality improvement programs. These initiatives relied on the timely peer review of physician cesarean delivery rates that are transparent, and placed in a framework of medical accountability (Gilbert et al., 2013). These initiatives also demonstrated a proportion of physicians make inappropriate medical decisions and improved training, and reinforcement by healthcare professionals is needed. Creating quality improvement standards that are locally owned and administratively reviewed establishes a system aimed at reducing inappropriate utilization as public health problem with the need for timely response. Quality improvement strategies also aim to change organizational behavior over time and create new norms that are acceptable and appropriate.

The inappropriate use of the cesarean delivery procedure is a proxy of the U.S health care system's performance and its quality, safety, and satisfaction. To intervene, health care reorganization is essential. Improving access to midwifery and hospitalist services show promise in reducing procedures among the privately insured having a low risk pregnancies and increasing the likelihood of vaginal delivery after cesarean delivery (Rosenstein, Nijagal, Nakagawa, Gregorich, & Kuppermann, 2015). These emerging models also show promise in reducing medical expenditures; at the population level, these models could result in significant cost savings for the nation. Creating public health programs that advance vaginal deliveries require a change in practice patterns to an expected scientific, societal, and economic norm. Introducing medical practice that are team based and driven by the systematic use of evidence and timely information, as well as having intrapartum plans for high risk women managed by interdisciplinary teams—these are the ever old, but new, expectations for national reform (Baldwin, Brodrick, Cowley, & Mason, 2010). This change aims to reverse inappropriate medical practice to an expected ethic aligned with reducing harm, improving solidarity in the statewide management of maternity patients among those having this responsibility, while maintaining local autonomy of provider practice and informed patient decision making. Otherwise, the significant reduction of the cesarean delivery procedure remains elusive, and instead, inappropriate utilization is a perpetuated normative standard despite health care reform.

Conclusion

There was significant variation in hospital ranks and association risks for the cesarean delivery procedure, according to risk adjustment models and statistical techniques. These differences may suggest bias in prevailing hospital profiling methods of the cesarean delivery procedure having applied logistic regression models versus the more recent endorsement of multilevel statistical techniques for U.S. health services research and policy making. Propensity score matching with stratification presented novel approaches for identifying approximately 10 – 20% of high risk maternity risk groups more likely representing appropriate utilization and other groups possibly representing a level of inappropriate utilization. Approaches from this research may be applied to health systems reform for the active monitoring of cesarean deliveries through statewide quality improvement programs for the reduction of inappropriate utilization and improved cost savings for the nation.

References

AABC American Association of Birth Centers. Uniform Data Set Perkiomenvilla, PA: AABC, 2010. Retrieved from http://www.birthcenters.org/?PDR

ACOG American College of Obstetrics and Gynecology (2010). ACOG Practice Bulletin No. 115: Vaginal birth after previous cesarean delivery. *Obstetrics & Gynecology*, *116*(2 Pt. 1), 450-463. http://journals.lww.com/greenjournal/pages/default.aspx

ACOG American College of Obstetrics and Gynecology (2015). FAQs: Cesarean birth. Retrieved from http://www.acog.org/Patients/FAQs/Cesarean-Birth-C-Section

ACOG Committee (2007). Opinion No. 386. Cesarean delivery on maternal request. *Obstetrics & Gynecology*, *110*(5), 1209-1212. Retrieved from http://journals.lww.com/greenjournal/pages/default.aspx

ACOG Committee (2006). Opinion No. 340. Mode of term singleton breech delivery. *Obstetrics & Gynecology*, *108*(1), 235-237. Retrieved from http://journals.lww.com/greenjournal/pages/default.aspx

ACOG Committee on Practice Bulletins. (2009). ACOG Practice Bulletin No. 107: Induction of labor. *Obstetrics & Gynecology*, *114*(2 Pt. 1), 386-397. Retrieved from http://journals.lww.com/greenjournal/pages/default.aspx

AHRQ National Quality Measures Clearinghouse. (2015). Agency for Healthcare Research and Quality, Washington DC. Retrieved from http://www.qualitymeasures.ahrq.gov.

Alexandrescu, R., Jen, M. H., Bottle, A., Jamare, B., Aylin, B (2011). Logistic Versus Hierarchical Modeling: An Analysis of a Statewide Inpatient Sample. *Journal of* *the American College of Surgeons*, 213(3), 392-401. doi: http://dx.doi.org/10.1016/j.jamcollsurg.2011.06.423

Allen, V., Baskett, T., & O'Connell, C. (2010). Contribution of select maternal groups to temporal trends in rates of caesarean section. *Journal of Obstetrics and Gynaecology Canada, 32*(7): 633-641. Retrieved from http://www.jogc.com/index_e.aspx

Aron, D. C., Harper, D. L., Shepardson, L. B., & Rosenthal, G. E. (1998). Impact of risk-adjusting cesarean delivery rates when reporting hospital performance. *Journal of the American Medical Association*, *279*(24), 1968-1972. doi:10.1001/jama.279.24.1968

Ash, A., Fienberg, S., Louis, T., Normand, S. L. T., Stukel, T., Stukel, T. A., … Utts, J. (2012). Statistical issues in assessing hospital performance. Commissioned by the Committee of Presidents of Statistical Societies. Submitted to CMS on January 27, 2012. Centers for Medicare & Medicaid Services, Washington DC. Retrieved from https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Statistical-Issues-in-Assessing-Hospital-Performance.pdf

Austin, P. C. (2008). Primer on statistical interpretation of methods report cards on propensity score matching in the cardiology literature from 2004-2006: A systematic review. *Circulation, 1*, 62-67. Retrieved from http://circ.ahajournals.org/

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*(3), 399-424. doi: 10.1080/00273171.2011.568786

Austin, P. C., Grootendorst, P., Normand, S. L. T., & Anderson G. (2007). Conditioning on the propensity score can result in biased estimation of common measures of treatment effects: A Monte Carlo study. *Statistics in Medicine*, *26*(4), 754-768. doi: 10.1002/sim.2618

Austin, P. C., & Mamdani, M. M. (2006). A comparison of propensity score methods: A case-study estimating the effectiveness of post-AMI statin use. *Statistics in Medicine*, *25*(12), 2084-2106. doi: 10.1002/sim.2328

Austin, P. C., Mamdani, M., Stukel, T., Anderson, G., & Tu., J. (2005). The use of propensity score for estimating treatment effects: Administrative versus clinical data. *Statistics in Medicine*, *24*(10), 1563 – 1578. doi: 10.1002/sim.2053

Austin, P. C., Tu, J. V., Alter, D. A. (2003). Comparing hierarchical modeling with traditional logistic regression analysis among patients hospitalized with acute myocardial infarction: should we be analyzing cardiovascular outcomes data differently? *American Heart Journal*, *145*(1), 27-35. Retrieved from http://www.ahjonline.com/article/S0002-8703(02)94732-0/abstract

Baicker, K., Buckles, K., & Chandra, A. (2006). Geographic variation in the appropriate use of cesarean delivery. *Health Affairs*, *25*(5), w355-w367. doi: 10.1377/hlthaff.25.w355

Bailit, J., Dooley, S. L., & Peaceman, A. N. (1999). Risk adjustment for interhospital comparison of primary cesarean rates. *Obstetrics & Gynecology*, *93*(6), 1025-1030. doi:10.1016/S0029-7844(98)00536-5.

Bailit, J., & Garrett, J. (2003). Comparison of risk adjustment methodologies for cesarean section rates. *Obstetrics & Gynecology*, *102*(1), 45-51. Retrieved from http://journals.lww.com/greenjournal/pages/default.aspx

Baldwin, J., Brodrick, A., Cowley, S., & Mason, N. (2010). An innovative toolkit to support normal birth. *Journal of Midwifery & Women's Health*, *55*(3), 270-272. doi: 10.1016/j.jmwh.2009.11.013

Barber, E. L., Lundsberg, L. S., Belanger, K., Pettker, C. M., Funai, E. F., & Illuzzi, J. L. (2011). Indications contributing to the increasing cesarean delivery rate. *Obstetrics & Gynecology*, *118*(1), 29-38. doi: 10.1097/AOG.0b013e31821e5f65

Bardach, N. S., Chien, A. T., & Dudley, R. A. (2010). Small numbers limit the use of the inpatient pediatric quality indicators for hospital comparison. *Academy of Pediatrics*, *10*(4), 266-73. doi:10.1016/j.acap.2010.04.025

Belitser, S. V., Martens, E. P., Pestman, W. R., Groenwold, R. H., de Boer, A., & Klungel, O. H. (2011). Measuring balance and model selection in propensity score methods. *Pharmacoepidemiol Drug Sa*fety, *20*(11), 1115-1129. doi: 10.1002/pds.2188

Bell, B., Ferron, J., & Kromrey, J. (2008). Cluster size in multilevel models: The impact of sparse data structures on point and interval estimates in two level models. Section on Survey Research Methods. *Joint Statistical Meeting,* 1122-1129.

Retrieved from http:// www.amstat.org/sections/srms/Proceedings/y2008/Files/300933.pdf

Berghella, V. (2007). *Maternal-fetal evidence-based guidelines*. Boca Raton, FL: CRC Press.

Berwick, D. M. (1994). Eleven worthy aims for clinical leadership of health system reform. *Journal of the American Medical Association, 72*(10), 797-802. doi:10.1001/jama.1994.03520100059034

Betran, A. P., Merialdi, M., Lauer, J. A., Ging-Shun, W., Thomas, J., Van Look, P.,. . . Wagner, M. (2007). Rates of caesarean section: Analysis of global, regional, and national estimates. *Paediatric & Perinatal Epidemiology*, *21*(2), 98-113. doi: 10.1111/j.1365-3016.2007.00786.x

Betran, A., Vindevoghel, N., Souza, J., Gulmesoglu, A., & Torloni, M. (2014). A systematic review of the Robson Classification for caesarean section: What works, doesn't work and how to improve it. *PLoS ONE, 9*(6), e97769. doi: 10.1371/journal.pone.0097769

Brennan, D., Robson, M., Murphy, M., & O'Herlihy, C. (2009). Comparative analysis of international cesarean delivery rates using 10 Group classification identifies significant variation in spontaneous labor. *American Journal of Obstetrics & Gynecology, 201*(3), e1-e8. doi: http://dx.doi.org/10.1016/j.ajog.2009.06.021

Burns, L. R., Geller, S. E., & Wholey, D. R. (1995). The effect of physician factors on the cesarean section decision. *Medical Care, 33*(4), 365-382. Retrieved from http://journals.lww.com/lwwmedicalcare/Abstract/1995/04000/The_Effect_of_Physician_Factors_on_the_Cesarean.4.aspx

Chaillet, N., & Dumont, A. (2007). Evidence-based strategies for reducing ceasarean section rates: A meta-analysis. *Birth, 34*(1), 53-64. doi: 10.1111/j.1523-536X.2006.00146.x

Cheschier, N., & Meints, L. (2009). Prospective study of coding practices for cesarean deliveries. *Obstetrics & Gynecology, 114* (2 Pt. 1), 217-223. doi: 10.1097/AOG.0b013e3181ad9533

Christianson, J. B., Volmar, K. M., Alexander, J., & Scanlon, D. P. (2010). A report care on provider report cards: Current status of the health care transparency movement. *Journal of General Internal Medicine, 25*(11), 1235-1241. doi: 10.1007/s11606-010-1438-2

Clancy, C. (2011). First, do not harm: Improving health quality and patient safety. Testimony of Carolyn Clancy before the Senate Health, Education, Labor and Pensions Committee. May 5, 2011. Retrieved from http://www.ahrq.gov/news/test050511.htm

Colorado Healthcare Affordability Act: Hospital Quality Improvement Payment ad hoc Committee (2010). Department of Health Care Policy & Financing. Denver, Colorado. Retrieved from http://www.chcpf.state.co.us.

Coonrod, D., Drachman, D., Hobson, P., & Manriquez, M. (2008). Nulliparous term singleton vertex cesarean delivery rates: Institutional and individual level predictors. *American Journal of Obstetrics & Gynecology*, *198*(6), 694.e1-e.8. doi: http://dx.doi.org/10.1016/j.ajog.2008.03.026

Costa, M., Cecatti, J., Souza, J., Milanez, H., & Gulmezoglu, M. (2010). Using a caesarean section classification system based on characteristics of the population as a way of monitoring obstetric patients. *Reproductive Health, 7*, 13. doi: 10.1186/1742-4755-7-13

Declerq, E., MacDorman, M., Osterman, M., Belanoff, C., & Iverson, R. (2015). Prepregnancy obesity and primary cesarean among otherwise low risk mothers in 38 States in 2012. *Birth, 42*(4), 309-318. doi: 10.1111/birt.12201

DeLong, E., Peterson, E., DeLong, D., Muhlbaier, L., Hackett, S., & Mark, D. (1997). Comparing risk adjustment methods for providing profiling. *Statistics in Medicine, 16*(23), 2645-2664. doi: 10.1002/(SICI)1097-0258(19971215)16:233.0.CO;2-D

Denk, C., Kruse, L., & Jain, N. (2006). Surveillance of cesarean section deliveries, New Jersey, 1999-2004. *Birth, 33*(3), 203-209. doi: 10.1111/j.1523-536X.2006.00105.x

D'Errigo, P., Tosti, M.E., Fusco, D., Perucci, C.A., & Seccareccia, F. (2007). Use of hierarchical models to evaluate performance of cardiac surgery centres in the Italian CABG outcome study. *BMC Medical Research Methodology, 7*, 29. doi:10.1186/1471-2288-7-29

DHHS U.S. Department of Health and Human Services. (2011). National Strategy for Quality Improvement in Health Care. Rockville, MD. Retrieved from http://www.ahrq.gov/workingforquality/

DiGiuseppe, D. L., Aron, D. C., Payne, S. M., Snow, R. J., Dierker, L., & Rosenthal, G. E. (2001). Risk adjusting cesarean delivery rates: A comparison of hospital profiles based on medical record and birth certificate data. *Health Services Research*, *36*(5), 959-977. Retrieved from http://www.hsr.org/hsr/abouthsr/journal.jsp

Dillon, W. P., Choate, J. W., Nusbaum, M. L., McCarthy, M. A., McCall, M., & Rosen, M. G. (1992). Obstetric care and cesarean birth rates: A program to monitor quality of care. *Obstetrics & Gynecology*, *80*(5), 731-737. Retrieved from http://journals.lww.com/greenjournal/pages/default.aspx

Donabedian, A. (2002). *An introduction in quality assurance in health care*. Oxford, England: Oxford University Press.

Donovan, E. F., Lannon, C., Bailit, J., Rose, B., Iams, J. D., & Byczkowski, T. (2010). A statewide initiative to reduce inappropriate scheduled births at 36(0/7) - 38(6/7) weeks' gestation. *American Journal of Obstetrics & Gynecology*, *202*(3), 243.e1-8. doi: 10.1016/j.ajog.2010.01.044

Ehrenthal, D. B., Hoffman, M. K., Jiang, X., & Ostrum, G. (2011). Neonatal outcomes after implementation of guidelines limiting elective delivery before 39 weeks of gestation. *Obstetrics & Gynecology*, *118*(5), 1047-1055. doi: 10.1097/AOG.0b013e3182319c58

Elixhauser, A., & Andrews, R. M. (2010). Profile of inpatient operating room procedures in US hospitals in 2007. *Archives of Surg*ery, *145*(12):1201-1208. doi: 10.1001/archsurg.2010.269

Emanuel, E., & Fuchs, V. (2008). The perfect storm of overutilization. *Journal of the American Medical Association*, *299*(23), 2789-2791. doi: 10.1001/jama.299.23.2789

Fantini, M., Stivinallo, E., Frammiartina, B., Barone, A., Fusco, D., Dallolio, L., . . . Perucci, C. (2006). Risk adjustment for inter-hospital comparison of primary cesarean section rates: Need, validity, and parsimony. *BMC Health Services Research*, *6*, 100. doi: 10.1186/1472-6963-6-100

Farley, D. O., & Battles, J. B. (2009). Evaluation of the AHRQ patient safety initiative: Framework and approach. *Health Services Research*, *44*(2 Pt. 2), 628-645. doi: 10.1111/j.1475-6773.2008.00931.x

Fisch, J. M., English, D., Pedaline, S., Brooks, K., & Simhan, H. N. (2009). Labor induction process improvement: A patient quality-of-care initiative. *Obstetrics & Gynecology*, *113*(4), 797-803. doi: 10.1097/AOG.0b013e31819c9e3d

Fisher, A., LaCoursiere, Y., Barnard, P., Bloebaum, L., & Varner, M. (2005). Differences between hospitals in cesarean rates for primigravidas with cephalic presentation. *Obstetrics & Gynecology*, *105*(4), 816-821. doi: 10.1097/01.AOG.0000156299.52668.e2

Flamm, B. L., Berwick, D. M., & Kabcenell, A. (1998). Reducing cesarean section rates safely: Lessons from a breakthrough series collaborative. *Birth*, *25*(2), 117-124. doi: 10.1046/j.1523-536x.1998.00117.x

Florica, M., Stephansson, O., & Nordstrom, L. (2006). Indications associated with increased cesarean section rates in a Swedish hospital. *International Journal of Gynecology & Obstetrics*, *92*(2), 181-185. doi: http://dx.doi.org/10.1016/j.ijgo.2005.10.016

Freeman, R. K., Cohen, A. W., Depp, R., Frigoletto, F. D., Hankins, G. D. V., Lieberman, E., . . . Sams, L. (2000). *Evaluation of cesarean delivery*. American College of Obstetrics and Gynecologists. Task Force on Cesarean Delivery Rates. Library of Congress: Washington, DC. Retrieved from http://images.ibsys.com/2005/0504/4450560.pdf

Fung, V., Schmittdiel, J., Fireman, B., Meer, A., Thomas, S., Snider, N., . . . Shelby, J.V. (2010). Meaningful variation in performance: A systematic literature review. *Medical Care*, *48*(2), 140-148. doi: 10.1097/MLR.0b013e3181bd4dc3

Gilbert, W. M., Bliss, M. C., Johnson, A., Farrell, W., Gregg, L., & Swanson, C. (2013). Improving recording accuracy, transparency, and performance for obstetric quality measures in a community hospital-based obstetrics department. *Joint Commission Journal on Quality and Patient Safety*, *39*(6), 258-266. Retrieved from http://www.ingentaconnect.com/con10t/jcaho/jcjqs/

Gilbert, S., Grobman, W., Landon, M., Spong, C., Rouse, D., Leveno, K. . . . Mercer, B. M. (2012). Elective repeat cesarean delivery compared with spontaneous trial of labor after a prior cesarean delivery: A propensity score analysis. *American Journal of Obstetrics & Gynecology*, *206*(4), 311.e1-e9. doi: 10.1016/j.ajog.2012.02.002

Giles, D., Austin, D., & Freymann, G. (2010). Enhanced data warehousing methods for creating and evaluating linked data for public health practice. Centers for Disease Control and Prevention, Atlanta, GA.

Glance, L. G., Dick, A. W., Osler, T. M., & Mukamel, D. B. (2006). Accuracy of hospital report cards based on administrative data. *Health Services Research*, *41* (4 Pt. 1), 1413-1437. doi: 10.1111/j.1475-6773.2006.00554.x

Glantz, J. (1999). Cesarean delivery risk adjustment for regional interhospital comparisons. *American Journal of Obstetrics & Gynecology*, *181*(6), 1425-1431. Retrieved from http://www.ajog.org/article/S0002-9378(99)70387-X/pdf

Goldman, L. E., Chu, P. W., Bacchetti, P., Kruger, .J, & Bindman, A. (2015). Effect of present-on-admission (POA) reporting accuracy on hospital performance assessments using risk-adjusted mortality. *Health Services Research*, *50*(3):922-38. doi: 10.1111/1475-6773

Gregory, K. D., Hackmeyer, P., Gold, L., Johnson, A. I., & Platt, L. D. (1999). Using the continuous quality improvement process to safely lower the cesarean section rate. *Joint Commission Journal on Quality & Patient Safety*, *15*(12), 619-629. Retrieved from http://www.ingentaconnect.com/content/jcaho/jcjqs/

Gregory, K., Korst, L., Gornbein, J., & Platt, L. (2002). Using administrative data to identify indications for elective primary cesarean delivery. *Health Services Research*, *37*(5), 1387-1400. doi: 10.1111/1475-6773.10762

Hamel, R., & Nairn, T. (2011). Ethical perspectives on health care reform: The individual mandate. *Journal of the Catholic Health Association of the United States*, *92*(4), 88-96. Retrieved from https://www.chausa.org/publications/health-progress/article/july-august-2011/ethical-perspectives-on-health-care-reform---the-individual-mandate

Handler, A. (1998). *Analytic methods in maternal child health*. University of Michigan Library: Ann Arbor, MI.

Hanley, J. A., McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*, 29-36.

Hofer, T., Hayward, R., Greenfield, S., Wagner, E., Kaplan, S., & Manning, W. G. (1999). The unreliability of individual physician 'report cards' for assessing the costs and quality of care of a chronic disease. *Journal of the American Medical Association*, *281*(22), 2098-2105. doi:10.1001/jama.281.22.2098

Homer, C., Kurinczuk, J., Spark P., Brocklehurst, P., & Knight, M. (2010). A novel use of a classification system to audit severe maternal morbidity. *Midwifery*, *26*(5), 532-536. doi: 10.1016/j.midw.2010.03.010

Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression*. New York, NY: John Wiley & Sons.

Houchens, R., Chu, B., & Steiner, C. (2007). Hierarchical modeling using H-CUP data. H-CUP Methods Series Report #2001-01. Agency for Healthcare Research and Quality, Washington DC. Retrieved from http://www.hcup-us.ahrq.gov/reports/methods.jsp.

Howell, S., Johnston, T., & Macleod, S. L. (2009). Trends and determinants of caesarean sections births in Queensland, 1997-2006. *Australian and New Zealand Journal of Obstetrics & Gynaecology*, *49*(6), 606-611. doi: 10.1111/j.1479-828X.2009.01100.x

Huang, I. C., Dominici, F., Frangakis, C., Diette, G.B., Damberg, C. L., & Wu, A. W. (2005). Is risk-adjustor selection more important than statistical approach for provider profiling? Asthma as an example. *Medical Decision Making*, *25* (1), 20–34. doi: 10.1177/0272989X04273138

Huang, I. C., Frangakis, C., Dominici, F., Diette, G., & Wu, A. (2005). Application of a propensity score approach for risk adjustment in profiling multiple physician groups on asthma care. *Health Services Research*, *40*(1), 253-278. doi: 10.1111/j.1475-6773.2005.00352.x

Huesch, M., & Doctor, J. M. (2015). Factors associated with increase cesearean risk among African American women: Evidence from California, 2010. *American Journal of Public Health*, *105*(5), 956-962. doi: 10.2105/AJPH.2014.302381

Iezzoni, L. (1997). The risks of risk adjustment. *Journal of the American Medical Association*, *278* (19), 1600-1607. doi:10.1001/jama.1997.03550190064046

Illinois Department of Public Health. Illinois Department of Public of Health Hospital Report Card (2012). Chicago, IL. Retrieved from http://www.healthcarereportcard.illinois.gov/

Imbens, G. (2000). The role of propensity score in estimating dose-response functions. *Biometrika*, *87*(3), 706-710. Retrieved from http://www.jstor.org/stable/2673642

James, B. C., & Savitz, L. A. (2011). How Intermountain trimmed health care costs through robust quality improvement. *Health Affairs*, *30*(6), 1185-1191. doi: 10.1377/hlthaff.2011.0358

John Paul II (May 1, 1991). *Centesimus Annus Encyclical Letter*. Apostolic Palace, Vatican City. http://w2.vatican.va/content/john-paul-ii/en/encyclicals/documents/hf_jp-ii_enc_01051991_centesimus-annus.html

Joseph, K., Young, D., Dodds, L., O'Connell, C., Allen, V., Chandra, S., . . . Allen A.C. (2003). Changes in maternal characteristics and obstetric practice and recent increase in primary cesarean delivery. *American Journal of Obstetrics & Gynecology*, *102*(4), 791-800. Retrieved from http://www.journals.elsevier.com/american-journal-of-obstetrics-and-gynecology/

Kabir, A., Steinmann, W., Myers, L., Khan, M., Herrera, E., & Joona, M. (2004). Unnecessary cesarean delivery in Louisiana: An analysis of birth certificate data. *American Journal of Obstetrics & Gynecology*, *190*(1), 10-19. doi: http://dx.doi.org/10.1016/j.ajog.2003.07.009

Kahn, E., Berg, C., & Callaghan, W. (2009). Cesarean delivery among women with low-risk pregnancies: A comparison of birth certificates and hospital discharge data. *Obstetrics & Gynecology*, *113*(1), 33-40. doi: 10.1097/AOG.0b013e318190bb33

Kazandjian, V. A., & Lied, T. R. (1998). Cesarean section rates: Effects of participation in a performance measurement project. *Joint Commission Journal on Quality Improvement & Patient Safety*, *24*(4), 187-196. Retrieved from http://www.ingentaconnect.com/content/jcaho/jcjqs/

Keeler, E. B., Park, R. E., Bell, R. M., Gifford, D.S., & Keesey, J. (1997). Adjusting cesarean delivery rates for case mix. *Health Services Research*, *32*(4), 511-528. Retrieved from http://www.hsr.org/

Kesmodel, U. S., & Jolving, L. R. (2011). Measuring and improving quality in obstetrics – the implementation of national indicators in Denmark. *Acta Obstetrics Gynecology Scandinavia*, *90*(4), 295-304. doi: 10.1111/j.1600-0412.2011.01078

Kilbourne, A. M., Fullerton, C., Dausey, D., Pincus, H. A., & Hermann, R. C. (2010). A framework for measuring quality and promoting accountability across silos: The case of mental disorders and co-occurring conditions. *Quality and Safety in Health Care*, *19*(2), 113-116. doi: 10.1136/qshc.2008.027706

Klazinga, N., Fischer, C., & Asbroek, A. (2011). Health services research related to performance indicators and benchmarking in Europe. *Journal of Health Services Research & Policy*, *16*(Suppl. 2), 38-47. doi: 10.1258/jhsrp.2011.011042

Kozhimannil, K., Law, M. & Virnig, B. (2013). Cesarean delivery rates vary 10-fold among US hospitals: Reducing variation may address quality, cost issues. *Health Affairs (Millwood)*, *32*(3), 527-535. doi: 10.1377/hlthaff.2012.1030

Kritchevsky, S. B., Braun, B. I., Gross, P. A., Newcomb, C. S., Kelleher, C. A., & Simmons, B. P. (1999). Definition and adjustment of cesarean section rates and assessments of hospital performance. *International Journal of Quality Health Care*, *11*(4), 283-291. doi: http://dx.doi.org/10.1093/intqhc/11.4.283 283-291

Krumholz, H., Brindis, R., Brush, J., Cohen, D., Epstein, J., Furie, K., … Normand, S. (2005). Standards for statistical models used for public reporting of health outcomes: An American Heart Association scientific statement from the quality of care and outcomes research interdisciplinary writing group: Cosponsored by the Council of Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. *Circulation*, *113*(3), 456-462. doi: 10.1161/CIRCULATIONAHA.105.170769

Kurth, T., Walker, A., Glynn, R., Chan, K., & Gaziano, J. (2005). Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of non-uniform effect. *American Journal of Epidemiology*, *163*(3), 262-270. doi: 10.1093/aje/kwj047

Kuss, O., Legler, T., & Börgermann, J. (2011). Treatments effects from randomized trials and propensity score analyses were similar in similar populations in an example from cardiac surgery. *Journal of Clinical Epidemiology*, *64*(10), 1076-1084. doi: 10.1016/j.jclinepi.2011.01.005

Lavis, J. N., & Anderson, G. M. (1996). Appropriateness in health care delivery: Definitions, measurement and policy implications. *Canadian Medical Association Journal*, *154*(3), 321-328. Retrieved from http://www.cmaj.ca/

Leape, L. L. (1992). Unnecessary surgery. *Annual Review of Public Health*, *13*, 363-83. doi: 10.1146/annurev.pu.13.050192.002051

Leung, K. M., Elashoff, R. M., Rees, K. S., Hasan, M. M., & Legorreta, A. P. (1998). Hospital and patient related characteristics determining maternity length of stay: A hierarchical linear model approach. *American Journal of Public Health*, *88*(3), 377-381. Retrieved from http://ajph.aphapublications.org/loi/ajph

Lin, H., & Xirasagar, S. (2004). Institutional factors in cesarean delivery rates: Policy and research implications. *Obstetrics & Gynecology*, *103*(1), 128-136. doi: 10.1097/01.AOG.0000102935.91389.53

Lydon-Rochelle, M., Holt, V., Cardenas, V., Nelson, J., Easterling, T., Gardella, C., . . . Callaghan, W. M. (2005). The reporting of pre-existing maternal medical conditions and complications of pregnancy on birth certificates and in hospital discharge data. *American Journal of Obstetrics and Gynecology*, *193*(1), 125-134. doi:: http://dx.doi.org/10.1016/j.ajog.2005.02.096

Lydon-Rochelle, M., Holt, V., Nelson, J., Cardenas, V., Gardella, C., Easterling, T. R., . . . Callaghan, W. M. (2005). Accuracy of reporting maternal in-hospital diagnoses and intrapartum procedures in Washington State linked birth records. *Paediatric Perinatal Epidemiology*, *19*(6), 460-471. doi: 10.1111/j.1365-3016.2005.00682.x

Main, E. K. (1999). Reducing cesarean birth rates with data-driven quality improvement activities. *Pediatrics*, *103*(1 Suppl E), 374-383. Retrieved from http://www.jpeds.com/

Main, E. K. (2009). New perinatal quality measures from the National Quality Forum, The Joint Commission and the Leapfrog Group. *Current Opinion on Obstetrics & Gynecology*, *21*(6), 532-540. doi: 10.1097/GCO.0b013e328332d1b0

Main, E. K., Bloomfield, L., & Hunt, G. (2004). Development of a large-scale obstetric quality-improvement program that focused on the nulliparous patient at term. *American Journal of Obstetrics & Gynecology*, *190*(6), 1747-1756. doi:: http://dx.doi.org/10.1016/j.ajog.2004.02.055

Main, E. K., Moore, D., Farrell, B., Schimmel, L. D., Altman, R. J., Abrahams, C., . . . Sterling, J. (2006). Is there a useful cesarean birth measure? Assessment of the nulliparous term singleton vertex cesarean birth rate as a tool for obstetric quality improvement. *American Journal of Obstetrics & Gynecology*, *194*(6), 1644-1651. doi: http://dx.doi.org/10.1016/j.ajog.2006.03.013

Martin, J. A., Hamilton, B. E., & Osterman, M. J. K. (2014). Births in the United States, 2013. *NCHS Data Brief*. No. 175, December 2015. Centers for Disease Control and Prevention, Washington DC. Retrieved from http://www.cdc.gov/nchs/data/databriefs/db175.htm

Martin, J., Wilson, E., Osterman, M., Saadi, E., Sutton, S., & Hamilton, B. E. (2013). Assessing the quality of medical and health data from the 2003 birth certificate revision: Results from two states. *National Vital Statistics Report*, *62*(2), 1-20. Retrieved from http://www.cdc.gov/nchs/products/nvsr.htm

Maso, G., Albreico, S., Monasta, L., Ronfani, L., Montico, M., Businelli, C., . . . Tamburlini, G. (2013). The application of the 10 Group Classification System (TGCS) in cesarean delivery case mix adjustment: A multicenter prospective study. *PLoS ONE*, *8*(6), e62364. doi: 10.1371/journal.pone.0062364

McCarthy, F., Rigg, L., Cady, L., & Cullinane, F. (2007). A new way of looking at Caesarean section births. *Australian and New Zealand Journal of Obstetrics & Gynecology*, *47*(4), 316-320. doi: 10.1111/j.1479-828X.2007.00753.x

Moineddin, R., Matheson, F. I., & Glazier, R. H. (2007). A simulation study of sample size for multilevel logistic regression models. *BMC Medical Research Methodology*, *16*, 7-34. doi:10.1186/1471-2288-7-34

Mukamel, D. B., Glance, L. G., Li, Y., Weimer, D. L., Spector, W. D., Zinn, J., . . . Mosqueda, L. (2008). Does risk adjustment of the CMS quality measures for nursing homes matter? *Medical Care*, *46*(5), 532-541. doi: 10.1097/MLR.0b013e31816099c5

Myers, S. A., & Gleicher, N. (1991). A successful program to reduce cesarean section rates: Friendly persuasion. *Quality Review Bulletin*, *7*(5), 162-166. Retrieved from http://www.unboundmedicine.com/medline/citation/2052351/A_successful_program_to_reduce_cesarean_section_rates:_friendly_persuasion.

NCHS National Center for Health Statistics. Healthy People 2000 Final Review. Hyattsville, Maryland: Public Health Service. 2001. Retrieved from http://origin.glb.cdc.gov/nchs/healthy_people/hp2000.htm

NCHS National Center for Health Statistics. Healthy People 2010 Final Review. Hyattsville, Maryland: Public Health Service. 2012. Retrieved from http://origin.glb.cdc.gov/nchs/healthy_people/hp2010.htm

National Quality Forum (2012). Retrieved from

http://www.healthcare.gov/law/resources/reports/quality03212011a.html.

New York Public Law S5153B-2011: Provides a premium reduction for physicians and licensed midwives who complete a risk management strategies course (2011). New York State Senate, Finance Committee, New York. Retrieved from http://open.nysenate.gov/legislation/api/1.0/html-print/bill/S5153B-2011.

Nicolay, C. R., Purkayastha, S., Greenhalgh, A., Benn, J., Chaturvedi, S., Phillips, N, . . . Darzi, A. (2012). Systematic review of the application of quality improvement methodologies from the manufacturing industry to surgical healthcare. *British Journal of Surgery*, *99*(3), 324-335. doi: 10.1002/bjs.7803

Normand, S. L. T., Glickman, M. E., & Gatsonis, C. A. (1997). Statistical methods for profiling providers of medical care: Issues and applications. *Journal of the American Statistical Association*, *92*(439), 803–814. doi: 10.2307/2965545

Normand, S. L. T., & Shahian, D. (2007). Statistical and clinical aspects of hospital outcomes profiling. *Statistical Science*, *22*(2), 206-226. Retrieved from http://www.jstor.org/stable/27645821

Orzag, P. (2008). The overuse, underuse, and misuse of health care. July 17, 2008. Testimony before the Committee on Finance, United States Senate, Office of Management and Budget, Washington, DC. Retrieved from https://www.cbo.gov/publication/41718

Oshiro, B. T., Henry, E., Wilson, J., Branch, D. W., & Varner, M. W. (2009). Decreasing elective deliveries before 39 weeks of gestation in an integrated health care system. *Obstetrics & Gynecology*, *113*(4), 804-811. doi: 10.1097/AOG.0b013e31819b5c8c

Paranjothy, S., Frost, C., & Thomas, J. (2005). How much variation in CS rates can be explained by case mix differences? *British Journal of Obstetrics & Gynecology, 112*(5), 658-666. doi: 10.1111/j.1471-0528.2005.00501.x

Patient Protection & Affordable Care Act, 111th U.S. Congress, Public Law 111-148, March 23, 2010. United States Congress, Washington DC.

Phillips, K. A., Morrison, K. R., Andersen, R., & Aday, L. A. (1998). Understanding the context of healthcare utilization: Assessing environmental and provider-related variables in the behavioral model of utilization. *Health Services Research*, *33*(3 Pt. 1), 571-596. Retrieved from http://www.hsr.org/hsr/abouthsr/journal.jsp

Profit, J., Typpo, K. V., Hysong, S. J., Woodard, L. D., Kallen, M. A., & Petersen, L. A. (2010). Improving benchmarking by using an explicit framework for the development of composite indicators: An example using pediatric quality of care. *Implementation Science*, *9*(5), 13. doi: 10.1186/1748-5908-5-13.

Qayad, M., & Zhang, H. (2009). Accuracy of data linkages. *Maternal Child Health Journal*, *13*(4), 531-538. doi: 10.1007/s10995-008-0377-6

Queenan, J. (2011). How to stop the relentless rise in cesarean deliveries. *Obstetrics & Gynecology*, 118(2 Pt. 1), 199-200. doi: 10.1097/AOG.0b013e3182266682

Racz, J. M., & Sedransk, J., (2010). Bayesian and frequentist methods for provider profiling using risk-adjusted assessments of medical outcomes. *Journal of the American Statistical Association, 105* (489), 48-58. doi: 10.1198/jasa.2010.ap07175

Rasmussen, O., Pedersen, B., Wilken-Jensen, C., & Vejerslev, L. (2000). Stratified rates of cesarean sections and spontaneous vaginal deliveries—Data from five labor wards in Denmark--1996. *Acta Obstetrics & Gynecology Scandinavia*, *79*(3), 227-231. doi: 10.1034/j.1600-0412.2000.079003227.x

Rassen, J. A., Glynn, R. J., Brookhart, M. A., & Schneeweiss, S. (2011). Covariate selection in high-dimensional propensity score analyses of treatment effects in small samples. *American Journal of Epidemiology*, *173*(12), 1404-1413. doi: 10.1093/aje/kwr001

RateMDs.com (2012). Retrieved from http://www.ratemds.com/.

Raudenbush, S. W., & Bryk, A. S. (1992). *Hierarchical linear models: Applications and data analysis methods* (1st ed.). Newbury Park, CA: Sage Publications.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage Publications.

Raudenbush S., Bryk A., Cheong, Y.F., Congdon R., Toit, M. (2004). *HLM 6: Hierarchical linear and nonlinear modeling.* Lincolnwood, IL: Scientific Software International, Inc.

Robson, M., Scudamore, I., & Walsh, S. (1996). Using the medical audit cycle to reduce cesarean section rates. *American Journal of Obstetrics & Gynecology*, *174*(1 Pt. 1), 199-206. doi: http://dx.doi.org/10.1016/S0002-9378(96)70394-0

Rosenbaum, P., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41-55. Retrieved from http://biomet.oxfordjournals.org/

Rosenstein, M., Nijagal, M., Nakagawa, S., Gregorich, S., & Kupperman, M. (2015). The association of expanded access to a collaborative midwifery and laborist model with cesarean delivery rates. *Obstetrics & Gynecology*, *126*(4), 716-723. doi: 10.1097/AOG.0000000000001032

Sage, W. M. (2009). *Solidarity: Unfashionable, but still American*. Connecting American Values with Health Reform. Garrison, NY: The Hastings Center.

Saha, D., Davila, A. A., Ver Halen, J. P., Jain, U. K., Hansen, N., Bethke, K., . . . Kim, J. Y. S. (2013). Post mastectomy reconstruction: A risk-stratified comparative analysis of outcomes. *Breast, 22*(6):1072-1080. Retrieved from http://dx.doi.org/10.1016/j.breast.2013.09.010

Scarella, A., Chamy, V., Sepulveda, M., & Belizan, J. (2010). Medical audit using the 10 Group Classification System and its impact on the cesarean section rate.

European Journal of Obstetrics & Gynecology Reproductive Biology, *154*(2), 136-140. doi: 10.1016/j.ejogrb.2010.09.005

Schifrin, B. S., & Cohen, W. R. (2012). The effect of malpractice claims on the use of caesarean section. *Best Practice Research in Clinical Obstetrics & Gynecology*, *27*(2), 269-283. doi: http://dx.doi.org/10.1016/j.bpobgyn.2012.10.004

Scott, J. R. (2011). Vaginal birth after cesarean delivery: A common-sense approach. *Obstetrics & Gynecology*, *118*(2 Pt. 1), 342-350. doi: 10.1097/AOG.0b013e3182245b39

Shahian, D. M., Edwards, F. H., Jacobs, J. P., Prager, R. L., Normand, S. L., Shewan, C., . . . Grover, F. L. (2011). Public reporting of cardiac surgery performance: Part 1 – history, rationale, consequences. *Annals of Thoracic Surgery*, *92*(3 Suppl.), S2-S11. doi: 10.1016/j.athoracsur.2011.06.100

Shahian, D., & Normand, L. (2008). Comparison of ‘risk-adjusted’ hospital outcomes. *Circulation*, *117*(15), 1955-1963. doi: 10.1161/CIRCULATIONAHA.107.747873

Shahian, D., Torchiana, D., Shemin, R., Rawn, J., & Normand, S. (2005). Massachusetts cardiac surgery report card: Implications of statistical methodology. *Annals of Thoracic Surgery*, *80*(6), 2106-2113. doi: http://dx.doi.org/10.1016/j.athoracsur.2005.06.078

Smeeth, L., Douglas, I., Hall, A. J., Hubbard, R., & Evans, S. (2009). Effect of statins on a wide range of health outcomes: A cohort study validated by comparison with randomized trials. *British Journal of Clinical Pharmacology*, *67*(1), 99-109. doi: 10.1111/j.1365-2125.2008.03308.x

Spalding, A. B. (1910). Cesarean section with report of nine cases. *California State Journal of Medicine, 8*(2), 49-52.

Spreeuwenberg, M. D., Bartak, A., Croon, M. A., Hagenaars, J. A., Busschback, J. J., Andrea, H., . . . Stijnen, T. (2010). The multiple propensity score as control for bias in the comparison of more than two treatment arms: An introduction from a case study in mental health. *Medical Care, 48*(2), 166-174. doi: 10.1097/MLR.0b013e3181c1328f

Srinivas, S. K., Fager, C., & Lorch, S. A. (2010). Evaluating risk-adjusted cesarean delivery rate as a measure of obstetric quality. *Obstetrics & Gynecology, 115*(5), 1007-1013. doi: 10.1097/AOG.0b013e3181d9f4b6

State of California, Office of Statewide Health and Planning (2012). State of California, Sacramento, California. Retrieved from http://www.oshpd.ca.gov/.

Stavrou, E. P., Ford, J. B., Shand, A. W., Morris, J. M., & Roberts, C. L. (2011). Epidemiology and trends for caesarean section births in New South Wales, Australia: A population based study. *BMC Pregnancy Childbirth, 11*, 8. doi: 10.1186/1471-2393-11-8

Steinbusch, P. J., Oostenbrink, J. B., Zuurbrier, J. J., & Schaepkens, F. J. (2007). The risk of upcoding in casemix systems: A comparative study. *Health Policy, 81*(2-3), 289-299. doi: http://dx.doi.org/10.1016/j.healthpol.2006.06.002

Steiner, P. M., Cook, T. D., Shadish, W. R., & Clark, M. H. (2010). The importance of covariate selection in controlling for selection bias in observational studies. *Psychology Methods, 15*(3), 250-267. doi: 10.1037/a0018719

Stivanello, E., Rucci, P., Carretta, E., Pieri, G., Seghieri, C., Nuti, S., . . . Fantini, M. P. (2011). Risk adjustment for inter-hospital comparison of caesarean delivery rates in low-risk deliveries. *PLoS One*. *6*(11), e28060. doi: 10.1371/journal.pone.0028060

Swensen, S., Kaplan, G., Meyer, G., Nelson, E., Hunt, G., Pryor, D. B., . . . Chassin, M. R. (2011). Controlling healthcare costs by removing waste: What American doctors can do now. *British Journal of Medical Quality & Safety*, *20*(6), 534-537. doi: 10.1136/bmjqs.2010.049213

The Joint Commission (2015, September). Joint Commission’s Specifications Manual for Joint Commission Quality Measures (Version 2012B). The Joint Commission, Washington DC. Retrieved from http://www.jointcommission.org/specifications_manual_for_national_hospital_inpatient_quality_measures.aspx

Thoemmes, F. (2012). Propensity score matching for SPSS. University of Tubingen, Tubingen, Germany. Retrieved from http://arxiv.org/ftp/arxiv/papers/1201/1201.6385.pdf

Thompson, J. (2009). 2009 Fact Sheet: Hospital Cost Controls. Washington State Department of Social and Health Services (2009), Spokane, WA.

Tita A., Landon M., Spong C., Lai Y., Leveno K., Varner, M. W., . . . Mercer, B. M. (2009). Timing of elective cesarean delivery at term and neonatal outcomes. *New England Journal of Medicine*, *360*(2), 111-120. doi: 10.1056/NEJMoa0803267

Torloni, M. R., Betran, A. P., Souza, J. P., Widmer, M., Allen, T., Gulmezoglu, M., . . . Cecatti, J. (2011). Classifications for cesarean section: A systematic review. *PLoS One*, *6*(1), e14566. doi: 10.1371/journal.pone.0014566

U.S. Agency for Healthcare Research and Quality (2015). HCUPnet, Healthcare Cost and Utilization Project. Rockville, MD: AHRQ. Retrieved from http://hcupnetahrq.gov/.

U.S. Department of Health and Human Services, Hospital Compare (2012). Retrieved from http://hospitalcompare.hhs.gov/.

Vermont Reimbursement for Avoidable Cesarean Sections. (2001). Vermont Bill H.392, 2011, VT Legislation 266011.1.

Watkins, K., Horvitz-Lennon, M., Caldarone, L. B., Shugarman, L. R., Smith, B., Mannle, T. E., … Pincus, H. A. (2011). Developing medical record-based performance indicators to measure the quality of mental healthcare. *Journal of Healthcare Quality*, *33*(1), 49-66. doi: 10.1111/j.1945-1474.2010.00128.x

Werner, R. M., & Asch, D. A. (2005). The unintended consequences of publicly reporting quality information. *Journal of the American Medical Association*, *293*(10), 1239-1244. doi:10.1001/jama.293.10.123.

West, S. G., Duan, N., Pequegnat, W., Gaist, P., Des Jarlais, D. C., Holtgrave, D., … Mullen, P. (2008). Alternatives to the randomized controlled trial. *American Journal of Public Health*, *98*(8), 1359-1366. doi: 10.2105/AJPH.2007.124446

World Health Organization (2015). WHO statement on cesarean section rates. Department of Reproductive Health, Geneva, Switzerland. Retrieved from http://www.who.int/mediacentre/news/releases/2015/caesarean-sections/en/

Ye, J., Betran, A., Vela, M., Souza, J., & Zhang, J. (2014). Searching for the optimal rate of medically necessary cesarean delivery. *Birth*, *41*(3), 237-244. doi: 10.1111/birt.12104

Zhang, J., Troendle, J., Reddy, U., Laughon, S., Branch D., Burkman, R., … Veldhuisen, P. (2010). Contemporary cesarean delivery practice in the United States. *American Journal of Obstetrics & Gynecology*, *203*(4), 326.e1-326.e.10. doi: 10.1016/j.ajog.2010.06.058

Appendix A: Logistic Regression Model for the Cesarean Delivery Outcome

The general multivariate logistic regression model included the following equation and varied according to entry of independent variables by model type (Hosmer & Lemeshow, 1998):

$$\ln (p/1\text{-}p) = b_0 + b_1X_1 + b_2X_2 + \ldots + bpXp$$

1. Whereas ln (p/1-p) was the expected logit of the odds that the cesarean delivery outcome are present.
2. Whereas $b_{0,}$ is the intercept.
3. Whereas $b_1X_1 + b_2X_2 + \ldots + bpXp$, represented selected risk adjustment variables.

Appendix B: Hierarchical Generalized Linear Model (HGLM) for the Cesarean Delivery Outcome

The HGLM model specifications included the following general Bernoulli model for the cesarean delivery outcome (Raudenbush & Byrk, 2002):

$$\phi_{ij} = \text{Prob}(\text{MOD}_{ij} = 1 \mid \beta_j) \qquad (1)$$

$$n_{ij} = \log [\, \phi_{ij} / (1 - \phi_{ij})] \qquad (2)$$

$$n_{ij} = \beta_{0j} + \beta_{1j} X_{1ij} + \beta_{2j} X_{2ij} + \ldots \beta_{pj} X_{pij} \qquad (3)$$

$$\beta_{qj} = \Upsilon_{q0} + \sum_{s=1}^{S_q} \Upsilon_{qs} Ws_j + u_{qj} \qquad (4)$$

1. Whereas (1) represented the probability of success for the Bernoulli distribution. MOD represented the method of delivery and whether a cesarean or vaginal event.
2. Whereas (2) represented the logit link for the binomial model.
3. Whereas (3) represented the Level 1 structural model and its patient level risk adjustment variables.
4. Whereas (4) represented the Level 2 model with random effects u_{qj} having a multivariate normal distribution with component means of zero and variance covariance matrix T.

Appendix C: Data Use Agreement

DATA USE AGREEMENT

This Data Use Agreement ("Agreement"), effective as of July 5, 2013 ("Effective Date"), is entered into by and between Denise Giles("Data Recipient") and Georgia Department of Public Health ("Data Provider"). The purpose of this Agreement is to provide Data Recipient with access to a Limited Data Set ("LDS") for use in research in accord with the HIPAA and FERPA Regulations.

1. Definitions. Unless otherwise specified in this Agreement, all capitalized terms used in this Agreement not otherwise defined have the meaning established for purposes of the "HIPAA Regulations" codified at Title 45 parts 160 through 164 of the United States Code of Federal Regulations, as amended from time to time.

2. Preparation of the LDS. Data Provider shall prepare and furnish to Data Recipient a LDS in accord with any applicable HIPAA or FERPA Regulations

3. Data Fields in the LDS. No direct identifiers such as names may be included in the Limited Data Set (LDS). In preparing the LDS, Data Provider shall include the **data fields specified as follows**, which are the minimum necessary to accomplish the research (contained in list of attachments below).

   a. 2a_Attachment_A_List of Birth Variables 1.4_DGiles _May_14_2013

   b. 2b_Attachment_B_2009_GDDS_DATA_SCHEMA_LDS further restricted by supplying the following surrogate fields

      i. A surrogate to identify the record(Public ID)

      ii. A surrogate to the longitudinal_id(ALONGID)

      iii. A surrogate to the facility id (AFACID)

      iv. A surrogate to provider(s) (APROVID(s))

   c. 2c_List of Fetal Death Variables 1.4

4. Responsibilities of Data Recipient. Data Recipient agrees to:

   a. Use or disclose the LDS only as permitted by this Agreement or as required by law;

   b. Use appropriate safeguards to prevent use or disclosure of the LDS other than as permitted by this Agreement or required by law;

c. Report to Data Provider any use or disclosure of the LDS of which it becomes aware that is not permitted by this Agreement or required by law;

d. Require any of its subcontractors or agents that receive or have access to the LDS to agree to the same restrictions and conditions on the use and/or disclosure of the LDS that apply to Data Recipient under this Agreement; and

e. Not use the information in the LDS to identify or contact the individuals who are data subjects.

5. Permitted Uses and Disclosures of the LDS. Data Recipient may use and/or disclose the LDS for its Research activities only.

6. Term and Termination.

a. Term. The term of this Agreement shall commence as of the Effective Date and shall continue for so long as Data Recipient retains the LDS, unless sooner terminated as set forth in this Agreement.

b. Termination by Data Recipient. Data Recipient may terminate this agreement at any time by notifying the Data Provider and returning or destroying the LDS.

c. Termination by Data Provider. Data Provider may terminate this agreement at any time by providing thirty (30) days prior written notice to Data Recipient.

d. For Breach. Data Provider shall provide written notice to Data Recipient within ten (10) days of any determination that Data Recipient has breached a material term of this Agreement. Data Provider shall afford Data Recipient an opportunity to cure said alleged material breach upon mutually agreeable terms. Failure to agree on mutually agreeable terms for cure within thirty (30) days shall be grounds for the immediate termination of this Agreement by Data Provider.

e. Effect of Termination. Sections 1, 4, 5, 6(e) and 7 of this Agreement shall survive any termination of this Agreement under subsections c or d.

7. Miscellaneous.

a. Change in Law. The parties agree to negotiate in good faith to amend this Agreement to comport with changes in federal law that materially alter either or both parties' obligations under this

Agreement. Provided however, that if the parties are unable to agree to mutually acceptable amendment(s) by the compliance date of the change in applicable law or regulations, either Party may terminate this Agreement as provided in section 6.

b. Construction of Terms. The terms of this Agreement shall be construed to give effect to applicable federal interpretative guidance regarding the HIPAA Regulations.

c. No Third Party Beneficiaries. Nothing in this Agreement shall confer upon any person other than the parties and their respective successors or assigns, any rights, remedies, obligations, or liabilities whatsoever.

d. Counterparts. This Agreement may be executed in one or more counterparts, each of which shall be deemed an original, but all of which together shall constitute one and the same instrument.

e. Headings. The headings and other captions in this Agreement are for convenience and reference only and shall not be used in interpreting, construing or enforcing any of the provisions of this Agreement.

IN WITNESS WHEREOF, each of the undersigned has caused this Agreement to be duly executed in its name and on its behalf.

DATA PROVIDER

Signed: David Austin

Print Name: DAVID AUSTIN

Print Title: DIRECTOR, DATA QUALITY AND ANALYSIS TEAM, OFFICE OF HEALTH INDICATORS FOR PLANNING, GEORGIA DEPARTMENT OF PUBLIC HEALTH

DATA RECIPIENT

Signed: Denise Fleles

Print Name: Denise FGles

Print Title: Ph.D Student

Health Scientist
Centers for Disease
Control, Malawi