Determinants of Intrapartum Quality of Care

Jennifer Bailit MD, MPH
David Aron, MD
Thomas Love, PhD

Case Western Reserve University
9/30/05-9/29/07
Project Officer: Judith Sangl

This project was supported by
grant number 5 R03 HS014352 from the Agency for Healthcare Research and Quality and the National Institute of Child Health and Development.

Structured Abstract

Purpose
To build primary cesarean delivery risk adjustment models and to determine the association between hospital quality and hospital structural and process of care factors.

Scope
Births in the State of California for 2003.

Methods
The inclusion and exclusion of race, ethnicity, splines, and interaction terms were explored to determine their effect on the risk adjustment model's ability to predict primary cesarean delivery. Models were used to rate hospitals as having primary cesarean rates that were above, within, or below expectations. Comparisons were made between risk-adjusted cesarean rates and hospital structural and process measures.

Results
In total, 371,648 patients were included. Race and ethnicity did not contribute meaningfully to predictive ability (C statistic .766 with race and ethnicity, .764 without). Splines and interaction terms also did not contribute substantially (C statistic .767 with interaction terms, .765 with splines). Hospital structural factors explained a small part of the variation in risk-adjusted primary cesarean rates (C statistic .768). The data on hospital induction and augmentation rates were of poor quality and were not usable.

Key words: quality of care, obstetrics, birth certificate, risk-adjusted cesarean rates.

Purpose

The purpose of this grant was threefold: to build primary cesarean delivery risk adjustment models and classify hospitals by quality as measured by their expected and actual primary cesarean delivery rates; to determine the association between hospital quality and hospital structural and process of care factors; and to develop and validate models of the relationships between hospital structural and process of care factors and risk-adjusted primary cesarean rates. Our two hypotheses were that Level-3 hospitals, rated as being able to provide the most complex perinatal care, will provide higher quality care than Level-1 and -2 hospitals and that low levels of labor induction at term will be associated with higher quality care.

Scope

Obstetrical quality indicators are an important part of the national health agenda. In part, this is due to the sheer number of deliveries each year. There were just over four million deliveries in the US in the year 2001, making delivery of an infant the second leading cause of hospitalization in the US.[1, 2] AHRQ has included several childbirth measures of quality in its Healthcare Cost and Utilization Project (HCUP).[3]

An ideal quality measure for inpatient obstetrics would encompass five major characteristics: 1) association with meaningful maternal and neonatal outcomes; 2) relation to outcomes that are influenced by physician/health system behaviors; 3) reliability and reproducibility; 4) inexpensive to apply on a large scale basis; and 5) acceptability to practicing obstetricians as a meaningful marker of quality.

Risk-adjusted primary cesarean rates meet many of the criteria for a good obstetrical quality marker. Risk-adjusted primary cesarean rates are particularly appealing because they are associated with both maternal and neonatal outcomes.[4] Hospitals that have risk-adjusted primary cesarean rates that are below expected have higher rates of poor maternal and neonatal outcomes.[4-7] Risk-adjusted cesarean rates do not provide a "target" cesarean rate. They do not pass judgment as to whether any particular cesarean was appropriate, and they do not attempt to assess the quality of surgical technique. The model simply predicts each patient's chance of a cesarean delivery given their personal risk factors in the hands of a typical provider. An institution's predicted rate is based solely on its case mix.

Having a quality marker for obstetrics allows us to explore the characteristics of hospitals with high quality of care and the processes of care that are more evident in these high-quality hospitals. The identification of these factors will help focus quality improvement efforts in the future. California is a large state with over 500,000 of the 4,000,000 births per year in the United States. It is diverse both in its population and in the variety of types of hospitals and hospital systems found there. Furthermore, California routinely prepares a data set each year of birth certificate data linked to maternal and infant discharge data (California PDD data). Thus, California is a good choice for studying a obstetric population for quality of care, with special attention to the issue of racial diversity. For our study, we used California PDD data for the year 2003, which was the most recent year available at the start of the grant.

Methods

After obtaining IRB approval from the MetroHealth Medical Center and the State of California Committee for the Protection of Human Subjects, we obtained 2003 California birth certificate data that had been linked to a hospital discharge data set for mothers and infants. All linkages are done by the State of California prior to release.[8]

The California data were then linked to the American Hospital Association annual survey for 2003 and the Accreditation Council for Graduate Medical Education's list of accredited obstetric residency programs in 2003. We limited the data set to women at risk for a primary cesarean delivery who delivered at a hospital having more than 50 deliveries per year. Additionally, we considered only viable deliveries – those births >24 weeks and >500 grams with no major anomalies. Last, we excluded patients with clearly mistaken entries (such as a vaginal delivery of a 15-lb infant).

A risk-adjustment model (model A) for primary cesarean delivery was created using multivariate logistic regression on the following predictor variables: maternal age, race, ethnicity, and medical conditions; gestational age; multiple births; insurance; nulliparity; complications of pregnancy; and the trimester in which prenatal care began. These variables have been previously identified as being important in a risk-adjustment model.[8] Clinically relevant categories of variables were created for most variables, including gestational age. Maternal age was expressed in years. We then created model A1 excluding race and ethnicity. We summarized the predictive validity and accuracy of the resulting models in several ways.

We calculated positive predictive value and negative predictive value for each model in predicting cesarean delivery across the entire sample. We compared C statistics (area under the receiving operating curve) to assess the discrimination of each model and Hosmer-Lemeshow tests to gauge model calibration.

To determine if the addition of interaction terms and splines would improve model performance, we built a series of more complex models to predict primary cesarean delivery. Model B adds product terms to model A to capture interaction effects between race/ethnicity, maternal age, and maternal medical conditions with the other risk factors. Model C adds cubic splines of gestational and maternal age to model B in order to account for nonlinear relationships between the risk factors and cesarean delivery.

In building these models, we made heavy use of methodologies for checking model fit and model assumptions.[10] These included detailed consideration of appropriate residuals as well as standard tests of goodness of fit and direct model comparisons.[11] To avoid overfitting, we performed extensive checks of model validation using both split-sample and bootstrap approaches.

We adapted a bootstrap resampling approach to the assessment of validation for each model we built.[10, 12] Our goal was to verify that our model's predicted values could accurately predict responses on subjects not included in building the model. Bootstrapping provides a method to estimate measures of statistical precision when no formula is otherwise available.

The crucial advantage of a bootstrap approach for model validation is that the bootstrap yields efficient and unbiased estimates of predictive accuracy. For each model, we made appropriate bootstrap point and interval estimates of the C statistic (area under the receiver operating characteristic curve) by sampling with a replacement to generate 200 replications of the entire data set.[10 13]

In a final validation step, we randomly split the data into a model development sample and an evaluation sample of 100,000 births at risk for primary cesarean delivery. We then developed the models described above on the development sample and checked C statistics and calibration measures on the validation sample. In light of our very large sample size, we anticipated (and observed) largely comparable results between the bootstrap and split-sample validation procedures.

The final model included maternal age, gestational age, multiple births, insurance, parity, start of prenatal care, complications of pregnancy, and medical condition. We proceeded to predict the risk of primary cesarean delivery for every woman in the data set. We created the predicted primary cesarean rate for the hospital by adding the individual probabilities for each woman delivering at that hospital. We then compared each hospital's predicted rate to the hospital's actual cesarean rate. To do this, we estimated a two-tailed 95% confidence interval around the ratio of the actual rate to the predicted rate. We then assigned the hospital to one of three groups based on the statistical tests implied by the confidence interval.[14] Specifically, if the confidence interval was entirely below 1, the hospital's primary cesarean rate is defined to fall below expectations. If the confidence interval's endpoints surrounded 1, the hospital's rate was defined as within expectations. Finally, if the confidence interval exceeded 1, the hospital's rate was defined as above expectations. Hospitals within expectations were categorized as good quality, and hospitals either below or above expectations were categorized as having poor quality.

Having assigned each hospital to a risk-adjusted group of above, within, or below expected rates, we added back to the data set women who were not at risk for a primary cesarean delivery. We also created a subset of the resulting data set that included only full-term gestations. These data sets will be referred to as the **complete** data set and the **term** data set, respectively.

Using the **complete** data set, we compared hospital structural factors with the hospital risk-adjusted categorizations, presence of a obstetrics residency, physician staffing arrangements (independent practice association, group practice without walls, open physician hospital organization, closed physician-hospital organization, management service organization, integrated salary model, equity model, or foundation), ratio of RNs to average daily hospital census,[15] ratio of LPNs to average daily hospital census,[15] organizational structure of the hospital (contract managed, member of an alliance, participate in a network), accreditation (Joint Commission, ACGME, medical school affiliation, nursing school affiliation, member of council of teaching hospitals, registered osteopathic hospital), and ownership of the hospital (government [non-federal], government [federal], non-government [not for profit], and investor-owned [for profit]). Hospital structural factors were then added to the final risk-adjustment model to see if they added any changes to the model's predictive ability.

We divided hospital structural factors into those that were modifiable and non-modifiable. Modifiable factors included physician staffing arrangements, RN to average daily census ratio, LPN to average daily census ratio, and hospital accreditation. We then added only modifiable structural factors to the final risk-adjustment model to assess the impact of modifiable factors on the models predictive ability.

Using the **term** data set, we examined the rates of labor induction at each hospital and by hospital risk-adjustment rating for women between the gestational ages of 37-40 weeks. We repeated these analyses with labor augmentation rather than labor induction. Data for labor induction and augmentation came from birth certificate records. We then added process factors, induction and augmentation, to the final risk-adjustment model. Lastly, we added structural factors and process factors to the final risk-adjustment model to look for changes in the model's predictive ability.

Results

Principal findings for Aim 1: To build primary cesarean delivery risk adjustment models and classify hospitals by quality as measured by their expected and actual primary cesarean delivery rates.

After cleaning and exclusions, there were a total of 382,566 deliveries in the data set to study Aim 1. Models with and without race/ethnicity show similar performance (Table I). The positive and negative predictive values are very close and the overall percent correct are similar. The C statistics for models with and without race and ethnicity are very close (0.7628 for model A with race and 0.7617 for model A1, without race and ethnicity) suggesting nearly identical levels of discrimination. As for calibration, full sample Brier scores (0.117 in each model) and Hosmer-Lemeshow test results are also very similar with and without race and ethnicity.

The odds ratios for all variables in the models A and A1 are shown in Table II. Nonetheless, the very modest differences in estimated odds ratios are associated with a highly statistically significant difference according to a likelihood-ratio test (chi-squared = 517 on 5 df, $p < 0.0001$). The very small P value in this comparison is mostly due to the enormous sample size available, as the two models are nearly indistinguishable using several metrics of predictive validity.

For more extensive validation, we based additional comparisons on split-sample and bootstrap-based assessments. In Figure 1, we present split-sample calibration plots based on the Hosmer-Lemeshow test for the two models. Each model appears to be generally and similarly well calibrated, except for a bump between predicted risks of approximately 0.5 to 0.8 when the sample sizes are relatively small.

The results of adding product terms and restricted cubic splines to model A (including race and ethnicity) are shown in Table III. Model A is the main effects model, Model B adds product terms to Model A, and Model C adds restricted cubic splines for maternal and gestational age to Model B. Model B shows a slightly higher C statistic than the other two approaches, though the are differences are small. Validation through bootstrapping or splitting the sample shows minimal differences in C statistics between the methods.

The Hosmer-Lemeshow plots, Brier scores, and pseudo R^2 statistics show that model A, B, and C are well calibrated and that there are minimal differences in calibration between the modes. The Brier scores are .1171 for model A, .1169 for model B, and .1171 for model C; the pseudo R^2 statistics are .1735, .1747, and .1731 for A, B, and C, respectively, for the entire sample of 371,468 observations (Hosmer-Lemeshow plots for these models are not shown here – split-sample P values exceed 0.2 for all three models).

Table I. Assessing the fit of model with and without race and ethnicity

Model	Description	Positive Predicted Value	Negative Predicted Value	% Classified Correctly	Split Sample C Statistic	Boot Strap Bias-Corrected C statistic (95% CI)
A	INCLUDES race and ethnicity	80.58%	85.18%	84.99 %	0.763	0.766 (0.764, 0.767)
A1	EXCLUDES race and ethnicity	80.58%	85.16%	84.97%	0.762	0.764 (0.763, 0.765).

Table II. Odds ratios (95% CI) for primary cesarean risk adjustment models with and without race and ethnicity, using the entire data set

Variable	With race and ethnicity	Without race and ethnicity
Maternal age	1.071 (1.069, 1.073)	1.067 (1.066, 1.069)
Gestational age		
24-30 wks vs. 37-40 wks	2.151 (1.923, 2.405)	2.195 (1.963, 2.454)
31-36 wks vs. 37-40 wks	1.234 (1.192, 1.278)	1.247 (1.205, 1.291)
41+ wks vs. 37-40 wks	1.264 (1.235, 1.295)	1.265 (1.235, 1.295)
Multiple births	14.023 (13.268, 14.820)	13.822 (13.080, 14.606)
Insurance	0.824 (0.805, 0.843)	0.776 (0.759, 0.792)
Nulliparity	5.048 (4.939, 5.159)	4.887 (4.783, 4.992)
Start of prenatal care		
First Trimester vs. No Care	1.435 (1.151, 1.790)	1.429 (1.146, 1.782)
Second Trimester vs. No Care	1.306 (1.045, 1.632)	1.302 (1.042, 1.626)
Third Trimester vs. No Care	1.279 (1.014, 1.614)	1.272 (1.009, 1.604)
Complications of pregnancy	45.097 (42.250, 48.137)	44.662 (41.846, 47.669)
Medical Conditions	1.058 (1.043, 1.074)	1.050 (1.034, 1.066)
Race		
(Black vs. White)	1.465 (1.404, 1.529)	
(Native Amer vs. White)	1.032 (0.898, 1.186)	
(Asian vs. White)	0.904 (0.876, 0.933)	
(Other vs. White)	1.067 (1.038, 1.097)	
Hispanic ethnicity	1.143 (1.116, 1.172)	

Table III Assessing the Fit of various Risk-Adjustment Models that all include race and ethnicity.

Model	Description	Positive Predicted Value	Negative Predicted Value	% Classified Correctly	Split Sample C Statistic	Boot Strap Bias-Corrected C statistic (95% CI)
A with race	Main Effects Only	80.58%	85.18%	84.99%	.7628	.7657 (.7640, .7672)
B with race	Adds product terms to A	79.69%	85.25%	85.01%	.7635	.7664 (.7649, .7682)
C with race	Adds age splines to B (maternal & gestational)	78.72%	85.28%	84.99%	.7621	.7647 (.7629, .7661)

* Model A contains the following predictors: maternal age and race, gestational age, indicators of multiple birth, insurance, mother's marital status, nulliparity, complications of pregnancy, maternal medical conditions, and the trimester in which prenatal care began.

Figure 1. Hosmer-Lemeshow Calibration Plot for model **including** race and ethnicity

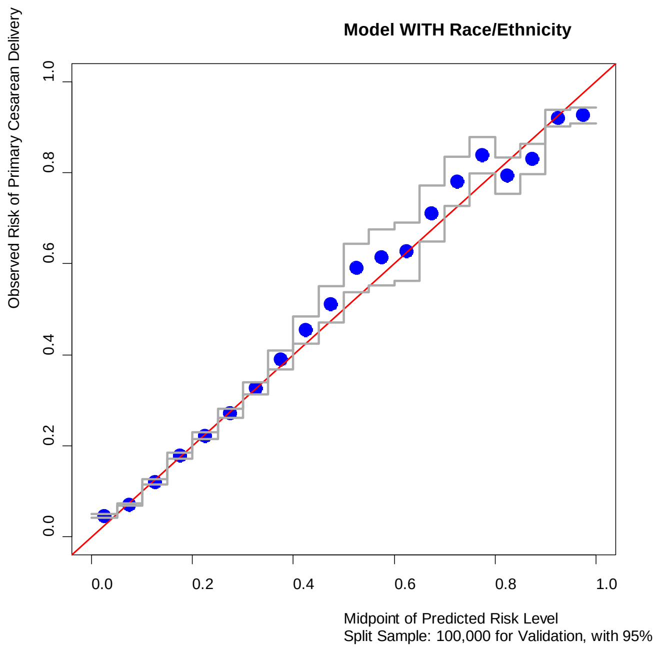
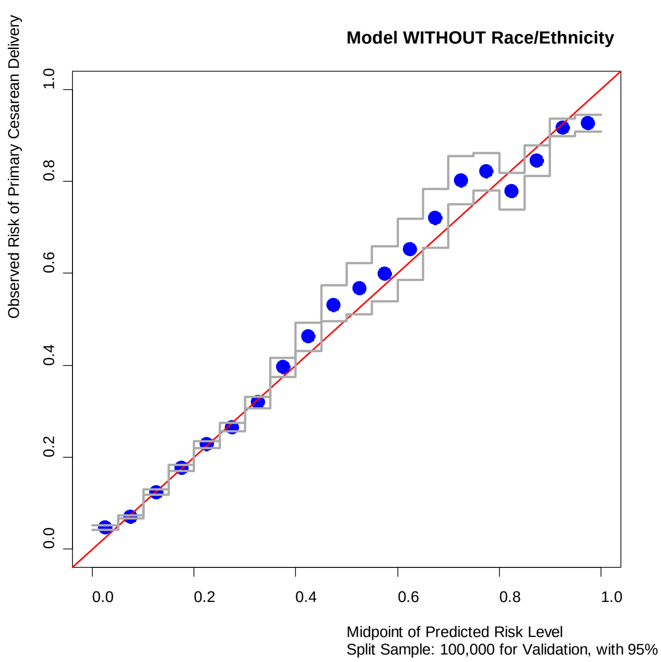


Figure 1a. Hosmer-Lemeshow Calibration Plot for model **not** including race and ethnicity



Discussion, significance, and implications of Aim 1. Our study suggests that removing race and ethnicity from risk-adjustment models has no substantial impact on predictive ability. Given the controversy that surrounds the inclusion of race and ethnicity in risk-adjustment models, it is important to understand the implications of leaving race and ethnicity in or out of the models. Regardless of any other concerns, these analyses suggest that race and ethnicity can be safely left out of primary cesarean rate risk-adjustment models. The impact of race and ethnicity on the predictive quality of the models is small enough that leaving them in cannot "explain away" substantial outcome differences due to discrimination. Nor does exclusion of race/ethnicity show any substantial evidence of important reductions in our ability to effectively adjust for differences in risk related to case mix.

Although this study design does not allow us to directly investigate the possibility, the fact that the models do not change substantially with and without race and ethnicity suggests that race's independent impact, if real, may in fact be thought of primarily as a marker for other model variables. Other authors have shown that medical conditions that impact pregnancy and hospitalizations for them vary by race.[16,17] Our model includes markers of both medical conditions and socioeconomics, and it appears that the incremental value of including race and ethnicity to such a model is still quite small.

Risk-adjustment models are difficult to understand for many practicing obstetricians. Our data show that model discrimination and calibration changes were minimal when product terms, polynomials, and splines were added. Keeping the models as simple as possible increases both the efficiency and clarity of the model. On the basis of these results, we believe that a main effects logistic regression model can safely be used to risk-adjust primary cesarean rates.

The strengths of our study are that it is based on the entire population in California and that the population is quite diverse. Furthermore, race is self-identified on birth certificates, suggesting that the data should be accurate. Though our results are consistent with the notion that race and ethnicity are markers for other processes that place a patient at risk for cesarean delivery, the study has limited ability to determine which other processes are responsible. Future work in this area may help to elucidate the mechanism through which race affects perinatal outcomes. Lastly, the value of statistical tests of significance is limited here, as the large data set renders these tests more or less useless in assessing the effectiveness of the two models. In this setting, even clinically unimportant differences between models appear to be statistically significant.

Conclusion aim 1: The mechanisms for how race and ethnicity affect perinatal outcomes may never be fully delineated. Despite this uncertainty, race and ethnicity show no substantial impact on the quality of risk-adjustment models for primary cesarean delivery.

Principal findings for Aim 2: To determine the association between hospital quality and hospital structural and process of care factors.

There were 371,643 patients available for analysis. This difference in the patients available for analysis from Aim 1 comes from using only data that had no missing variables for the key variable in the model.

One of our key hypotheses was that Level-3 hospitals, rated as being able to provide the most complex perinatal care, will provide higher quality care than Level-1 and -2 hospitals. Higher quality of care was considered to be a risk-adjusted primary cesarean rates (RAPCR) rating of "within." Lower quality of care was considered a RAPCR rating of "below." Table 4 shows the comparison between RAPCR and the level of perinatal care. Perinatal care level information was available for 258,509 of the 371,643 births with a RAPCR rating.

Table 4. Perinatal care level vs. RAPCR rating

OB_LEV = 1	Provides services for uncomplicated maternity and newborn cases
OB_LEV = 2	Provides service for all uncomplicated and most complicated cases
OB_LEV = 3	Provides services for all serious illnesses and abnormalities

Pearson χ^2 = 8187.12 on 4 df, so $p < 0.0001$	RAPCR Observed *Above* Expectations	RAPCR Observed *Within* Expectations	RAPCR Observed *Below* Expectations	Total
Basic OB Services [OB_LEV = 1]	26,349 [41% of Basic] [27% of Above]	18,875 [30% of Basic] [30% of Within]	18,282 [29% of Basic] [19% of Below]	63,506 [25% of Total]
Moderate OB Services [OB_LEV = 2]	41,901 [35% of Moderate] [43% of Above]	33,847 [28% of Moderate] [54% of Within]	44,248 [37% of Moderate] [46% of Below]	119,996 [47% of Total]
Tertiary OB Services [OB_LEV = 3]	29,417 [40% of Tertiary] [30% of Above]	10,281 [14% of Tertiary] [16% of Within]	34,309 [46% of Tertiary] [35% of Below]	74,007 [29% of Total]
Total	97,667 [38% of Total]	63,003 [24% of Total]	96,839 [38% of Total]	257,509

The majority of patients deliver at Level-2 hospitals (119,996) followed by tertiary care (74,007) and basic care (63,506) hospitals. Thirty percent of deliveries at basic-level hospitals occurred at hospitals with a higher quality of care as compared to 28% of deliveries in Level-2 hospitals. Patients delivering at tertiary care hospitals were the least likely (14%; p< .0001) to fall in the higher quality of care category. Thus, our hypothesis did not hold true. Quality of care appears to be higher in basic and Level-2 hospitals than in tertiary hospitals. This finding may be due to actual differences in quality of care or it might be explained by differing intensity of coding for problems between hospital types. More information on intensity of coding by hospital type needs to be known before this can be fully understood.

Tables 5-13 show the individual comparison of structural factors compared to RAPCR rating. All factors demonstrated statistical differences. Table 5. Joint Commission-accredited hospitals appear to have higher quality than nonaccredited hospitals (25% vs. 10% in the within category); however, most accredited hospitals (42%) were in the lower-quality below rating group. Table 6. Group practice without walls, despite having only 2% of the total population, had the most patients delivering in hospitals rated to be of high quality (81%). Table 7. Hospital district or authority control had the highest percentage (28%) of patients delivering in hospitals considered to be high quality. Table 8. Contract-managed hospitals had more patients delivering at high-quality hospitals (38%). Table 9. Network had information missing on too many hospitals to have any meaningful analysis.

Table 10 shows inductions vs. RAPCR rating. Patients without inductions (26%) were more likely to deliver at hospitals considered high quality. However, please see the discussion on Aim 3 for further information on the induction data. Table 11 shows that patients without augmentation (25%) were more likely to deliver in high-quality hospitals. Table 12's raw means and medians would suggested that higher RN/patient and LPN/patient ratios are more likely to be associated with higher quality hospitals, but this did not reach statistical significance with the exception of the mean LPN/patient ratio. However, the median LPN/patient ratio was not significant, suggesting that this finding requires careful interpretation .

Table 5. The Joint Commission (TJC) Accreditation vs. RAPCR rating

Pearson $\chi^2 = 650.28$ on 2 df, so $p < 0.0001$	RAPCR Observed *Above* Expectations	RAPCR Observed *Within* Expectations	RAPCR Observed *Below* Expectations	Total
JCAHO Accredited	120,317 [33% of TJC] [98.4% of Above]	91,795 [25% of TJC] [99.3% of Within]	154,027 [42% of TJC] [98.1% of Below]	366,139 [98.5% of Total]
Not *JCAHO* Accredited	2014 [37% of not TJC] [1.7% of Above]	570 [10% of not TJC] [0.6% of Within]	2920 [53% of not TJC] [1.9% of Below]	5504 [1.5% of Total]
Total	122,331 [33% of Total]	92,365 [25% of Total]	156,947 [42% of Total]	371,643

Table 6. Type of Care vs. RAPCR rating

	RAPCR Observed *Above* Expectations	RAPCR Observed *Within* Expectations	RAPCR Observed *Below* Expectations	Total
IPA: Independent Practice Association (Hospital)	45,967 [42% of IPA] [38% of Above]	26,394 [24% of IPA] [30% of Within]	37,024 [34% of IPA] [24% of Below]	109,385 [25% of Total]
GPWWW: Group Practice Without Walls (Hospital)	1250 [19% of GPWWW] [1% of Above]	5253 [81% of GPWWW] [6% of Within]	0	6503 [2% of Total]
OPHO: Open Physician-Hospital Organization (Hospital)	1457 [17% of OPHO] [1% of Above]	400 [5% of OPHO] [0.5% of Within]	6488 [78% of OPHO] [4% of Below]	8345 [2% of Total]
CPHO: Closed Physician-Hospital Organization (Hospital)	503 [7% of CPHO] [0.4% of Above]	2298 [32% of CPHO] [3% of Within]	4471 [61% of CPHO] [3% of Below]	7272 [2% of Total]
MSO: Management Service Organization (Hospital)	7261 [23% of MSO] [6% of Above]	11,166 [35% of MSO] [13% of Within]	13,486 [42% of MSO] [9% of Below]	31,913 [9% of Total]
ISM: Integrated Salary Model (Hospital)	7086 [28% of ISM] [6% of Above]	2514 [10% of ISM] [3% of Within]	15,564 [62% of ISM] [10% of Below]	25,164 [7% of Total]
Foundation (Hospital)	10,945 [49% of Found.] [9% of Above]	166 [0.7% of Found.] [0.2% of Within]	11,295 [50% of Found.] [7% of Below]	22,406 [6% of Total]
Equity model (Hospital)	0	0	0	0
None of These [N/A]	47,817 [31% of N/A] [39% of Above]	40,736 [26% of N/A] [46% of Within]	66,819 [43% of N/A] [43% of Below]	155,372 [42% of Total]
Total	122,286 [33% of Total]	88,927 [24% of Total]	155,147 [42% of Total]	366,360

Table 7. Locus of Control vs. RAPCR rating

	RAPCR Observed *Above* Expectations	RAPCR Observed *Within* Expectations	RAPCR Observed *Below* Expectations	Total
CONTROL = 12: State	1453 [20% of State] [1% of Above]	0	5804 [80% of State] [4% of Below]	7257 [2% of Total]
CONTROL = 13: County	2405 [13% of County] [2% of Above]	4067 [22% of County] [4% of Within]	11977 [65% of County] [8% of Below]	18,449 [5% of Total]
CONTROL = 14: City	914 [100% of City] [0.8% of Above]	0	0	914 [0.3% of Total]
CONTROL = 15: City-County	0	0	777 [100% of City-County] [0.5% of Below]	777 [0.2% of Total]
CONTROL = 16: Hospital district or authority	7216 [25% of Hospital] [6% of Above]	8092 [28% of Hospital] [9% of Within]	14,013 [48% of Hospital] [9% of Below]	29,321 [8% of Total]
CONTROL = 21: Church-operated	17,962 [33% of Church] [15% of Above]	12,788 [24% of Church] [14% of Within]	23,474 [43% of Church] [15% of Below]	54,224 [15% of Total]
CONTROL = 23: Other Non-government, not for profit	54,586 [29% of Other nfp] [45% of Above]	50,965 [27% of Other nfp] [55% of Within]	85,779 [45% of Other nfp] [55% of Below]	191,330 [51% of Total]
CONTROL = 31: Individual investor-owned (for profit)	829 [100% of Indiv.] [0.7% of Above]	0	0	829 [0.2% of Total]
CONTROL = 32: Partnership investor-owned (for profit)	7601 [100% of Partner] [6% of Above]	0	0	7601 [2% of Total]
CONTROL = 33: Corporation investor-owned (for profit)	29,365 [48% of Corp] [24% of Above]	16,453 [27% of Corp] [18% of Within]	15,123 [25% of Corp] [10% of Below]	60,941 [16% of Total]
GOVERNMENT, FEDERAL (codes 41-48)	0	0	0	0
Total	122,331 [33% of Total]	92,365 [24% of Total]	156,947 [42% of Total]	366,360

Table 8. Contract_Managed vs. RAPCR rating

Pearson χ^2 = 1042.78 on 2 df, so $p < 0.0001$	RAPCR Observed *Above* Expectations	RAPCR Observed *Within* Expectations	RAPCR Observed *Below* Expectations	Total
Hospital is contract-managed	1588 [21% of Contract] [1% of Above]	2897 [38% of Contract] [3% of Within]	3141 [41% of Contract] [2% of Below]	7626 [2% of Total]
Not contract-managed	96079 [37% of not Contract] [79% of Above]	64905 [25% of not Contract] [70% of Within]	99258 [38% of not Contract] [63% of Below]	260242 [70% of Total]
Missing information on contract_managed	24,664 [24% of Missing] [20% of Above]	24,563 [24% of Missing] [27% of Within]	54,548 [53% of Missing] [35% of Below]	103,775 [28% of Total]
Total	122,331 [33% of Total]	92365 [25% of Total]	156,947 [42% of Total]	371643

Table 9. Network vs. RAPCR Rating

	RAPCR Observed *Above* Expectations	RAPCR Observed *Within* Expectations	RAPCR Observed *Below* Expectations	Total
Hospital is a participant in a network	13,523 [33% of Total]	8459 [25% of Total]	16,442 [42% of Total]	38424

Information is MISSING on Network for the remaining 333,219 births (90% of the total of 371,643 births)

Table 10. Induction vs. RAPCR Rating

Pearson χ^2 = 4348.58 on 2 df, so $p < 0.0001$	RAPCR Observed *Above* Expectations	RAPCR Observed *Within* Expectations	RAPCR Observed *Below* Expectations	Total
Induction = 1	8881 [22% of Ones] [7% of Above]	8725 [21% of Ones] [9% of Within]	23,306 [57% of Ones] [15% of Below]	40,912 [11% of Total]
Induction = 0	113,450 [34% of Zeros] [93% of Above]	83,640 [25% of Zeros] [91% of Within]	133,641 [40% of Zeros] [85% of Below]	330,731 [89% of Total]
Total	122,331 [33% of Total]	92,365 [25% of Total]	156,947 [42% of Total]	371,643

Table 11. Augment vs. RAPCR Rating

Pearson χ^2 = 13410.45 on 2 df, so $p < 0.0001$	RAPCR Observed *Above* Expectations	RAPCR Observed *Within* Expectations	RAPCR Observed *Below* Expectations	Total
Augment = 1	5439 [13% of Ones] [4% of Above]	7696 [19% of Ones] [8% of Within]	27950 [68% of Ones] [18% of Below]	41,085 [11% of Total]
Augment = 0	116,892 [35% of Zeros] [96% of Above]	84,669 [26% of Zeros] [92% of Within]	128,997 [39% of Zeros] [82% of Below]	330,558 [89% of Total]
Total	122,331 [33% of Total]	92,365 [25% of Total]	156,947 [42% of Total]	371,643

Table 12. Descriptive Statistics and Hypothesis Test Results for Key Ratios (at the Obstetric Unit level) by RAPCR Rating Status

RN Ratio (FTE_RNs/ADC)	RAPCR Observed *Above* Expectations	RAPCR Observed *Within* Expectations	RAPCR Observed *Below* Expectations	**p value**
N	75	88	102	
Mean	95.56	116.40	99.68	**0.382**
Std Dev	116.41	105.86	91.02	
Median	65.06	82.82	71.23	**0.075**
(Minimum, Maximum)	(15.43, 844.71)	(12.60, 619.76)	(10.16, 537.33)	

ANOVA F test comparing the mean RN Ratio across three levels of Outlier gives F = 0.97 on 2 and 262 df, for $p = 0.382$
Kruskal-Wallis Rank Sum test comparing the median RN Ratio across Outlier levels gives $\chi^2 = 5.18$ on 2 df, for $p = 0.075$

LPN Ratio (FTE_LPNs/ADC)	RAPCR Observed *Above* Expectations	RAPCR Observed *Within* Expectations	RAPCR Observed *Below* Expectations	**ANOVA p value**
N	75	88	102	
Mean	15.06	20.21	11.53	**0.031**
Std Dev	28.71	25.19	13.00	
Median	8.00	12.18	7.48	**0.029**
(Minimum, Maximum)	(0.68, 227.42)	(0.41, 127.13)	(0.34, 90.79)	

ANOVA F test comparing the mean LPN Ratio across three levels of Outlier gives F = 3.52 on 2 and 262 df, for $p = 0.031$
Kruskal-Wallis Rank Sum test comparing the median LPN Ratio across Outlier levels gives $\chi^2 = 7.09$ on 2 df, for $p = 0.029$

Table 13. Descriptive Statistics for Raw Data Elements (at the Obstetric Unit level) by RAPCR Rating Status

FTE_RNs (# of FTE RNs)	RAPCR Observed *Above* Expectations	RAPCR Observed *Within* Expectations	RAPCR Observed *Below* Expectations
N	75	88	102
Mean	315.43	232.47	353.68
Std Dev	247.55	262.62	342.95
Median	234	161	238
(Minimum, Maximum)	(45, 1305)	(24, 1924)	(29, 1723)

FTE_LPNs (# of FTE LPNs)	RAPCR Observed *Above* Expectations	RAPCR Observed *Within* Expectations	RAPCR Observed *Below* Expectations
N	75	88	102
Mean	36.60	28.98	35.79
Std Dev	26.46	28.51	31.48
Median	29	20	27.5
(Minimum, Maximum)	(7, 139)	(1, 186)	(1, 193)

ADC (Average Daily Census)	RAPCR Observed *Above* Expectations	RAPCR Observed *Within* Expectations	RAPCR Observed *Below* Expectations
N	75	88	102
Mean	4.46	2.87	4.21
Std Dev	2.90	2.47	2.79
Median	4.02	1.92	3.62
(Minimum, Maximum)	(0.12, 13.14)	(0.06, 10.77)	(0.17, 14.59)

Discussion, significance, and implications Aim 2. Though most structural factors showed statistical significance, the most striking aspect of the data is that, even for the types of hospitals that had the highest number of patients delivering at high-quality hospitals, most of the patients in these categories still delivered at hospitals that were of lower quality. For example, patients delivering at The Joint Commission-accredited hospitals were more likely to deliver at hospitals of high-quality than at nonaccredited hospitals. However, most patients delivering at accredited hospitals were still in hospitals of lower quality.

Conclusions Aim 2: These findings suggest that all structural types of hospitals need improvement toward higher quality.

Principal findings for Aim 3: To develop and validate models of the relationships between hospital structural and process of care factors and risk-adjusted primary cesarean rates

Table 14 shows the predictive ability of the primary cesarean risk-adjustment model with and without hospital structural factors. The differences in predictive ability between the models are minimal. High C statistics indicate better model discrimination. Lower Brier scores indicate better calibration. Our results suggest that hospital structural factors explain only a small part of the differences in risk adjusted cesarean rates. Table 15 shows the predictive ability of the primary cesarean risk-adjustment model with and without process factors. The impact of process factors on the predictive ability of the model is minimal. Table 16 shows the effect of both structural and process factors on the predictive ability of the model. Once again, the predictive impact is minimal.

Table 14. Predictive ability of a primary cesarean risk-adjustment model with and without hospital structural factors.

Model	C statistic (95% CI)	Brier score	Positive predictive value (%)	Negative predictive value (%)
Maternal Factors only	0.764 (0.763, 0.765)	0.1141	80.58	85.16
Maternal factors + all hospital structural factors	0.768 (0.766,0.770)	0.1194	79.47	84.77
Maternal factors + modifiable hospital structural factors	0.765 (0.763,0.767)	0.1171	80.63	85.16

Table 15. Predictive ability of a primary cesarean risk-adjustment model with and without process factors.

Model	C statistic (95% CI)	Brier score	Positive predictive value (%)	Negative predictive value (%)
Maternal Factors only	0.764 (0.763, 0.765)	0.1141	80.58	85.16
Maternal Factors + induction and augmentation	0.766 (.764, 0.767)	0.1168	80.65	85.20

Table 16. Predictive ability of a primary cesarean risk-adjustment model with structural and process factors.

Model	C statistic (95% CI)	Brier score	Positive predictive value (%)	Negative predictive value (%)
Maternal Factors only	0.764 (0.763, 0.765)	0.1141	80.58	85.16
Maternal Factors + all process and structural factors	0.770 (.768, 0.772)	0.1190	79.66	84.82

Discussion, significance, and implications Aim 3. Hospital structural factors appear to have little impact on the predictive ability of our risk-adjustment model, suggesting that the hospital structure itself does not have a direct impact on the outcomes. Process factors are more difficult to interpret. Since this grant was funded, more information about the quality of induction data has been published. Romano et al. found that, when compared to the gold standard of a medical record, induction of labor in discharge data were only 45% sensitive and 88% specific.[18] We used birth certificate data rather than discharge data, but, looking at hospital rates of induction of labor, we believe the sensitivity of birth certificate data is likely to be similarly flawed. Of the 219 hospitals for which we have data, 29 (13%) had an induction rate of <1%, and 111 (50%) hospitals had a rate <10%. The overall induction rate average in our data set was 12.5%. Given estimates of induction rates nationally between 21.2% (from national birth certificate data) and 38.8%[19](from a well done single center study), a 12.5% induction rate and having 50% of hospitals with induction rates of <10% is not believable. Thus, we are unable to adequately evaluate our second hypothesis that low levels of labor induction at term will be associated with higher-quality care. Future studies on the impact of induction of labor will need to focus on other data sources or wait for improvement in the quality of birth certificate and hospital discharge data.

Conclusion for Aim 3: Hospital structural factors have little impact on the ability to predict primary cesarean delivery. Induction is not accurately recorded in birth certificate data sets.

Grant Bibliography

Bailit JL, Love TE. The role of race in cesarean rate case mix adjustment. American Journal of Obstetrics and Gynecology. In press.
Please note, this paper was originally two manuscripts and they were combined at the request of the journal editors.

Bailit JL, Love T, Performance of simple vs. complex models for risk adjusting primary cesarean delivery rates. Poster Presentation. Society for Medical Decision Making. Boston, MA, October 16, 2006.

Bailit, JL, Love, The role of race in predicting primary cesarean delivery rates. Poster presentation. 27th Annual Meeting of the Society of Maternal-Fetal Medicine. San Francisco, CA. February 10, 2007

Additionally, another abstract and another manuscript have been submitted. Another manuscript is close to completion and will be submitted shortly.

1. Martin J, Hamilton B, Ventura S, Manacker F, Park M, Sutton P. Births: Final data for 2001. National vital statistics reports; vol 51 no2. Hyattsville, MD: National Center for Health Statistics, 2002.
2. Hall M, Owings M. 2000 National hospital discharge survey. Advance data from vital and health statistics; no 329. Hyattsville, MD: National center for health statistics, 2002.
3. UCSF-Stanford evidence-based practice center. Refinement of the HCUP quality indicators technical review number 4. Rockville, MD: Agency for Healthcare Research and Quality, 2001.
4. Bailit J, Love T, Dawson N. Quality of obstetric care and risk-adjusted primary cesarean delivery rates. American Journal of Obstetrics and Gynecology 2006;194:402-07.
5. Bailit J, Garrett J, Miller W, McMahon M, Cefalo R. Hospital primary cesarean delivery rates and the risk of poor neonatal outcomes. Am J Obstet Gynecol 2002;187:721-27.
6. Gould JB, Danielsen B, Korst LM, Phibbs R, Chance K, Main E, et al. Cesarean delivery rates and neonatal morbidity in a low-risk population. Obstetrics & Gynecology 2004;104:11-19.
7. Li T, Rhoads GG, Smulian J, Demissie K, Wartenberg D, Kruse L. Physician cesarean delivery rates and risk-adjusted perinatal outcomes. Obstetrics & Gynecology 2003;101:1204-12.
8. Danielsen B. Probabilistic record linkages for Generating a Comprehensive Epidemiological Research File on Maternal and Infant Health, 2000. Available at http://health-info-solutions.com/links.html. Accessed March 8, 2007.
9. Bailit J, Shulkin J, Dawson N. Risk-adjusted cesarean rates:What risk factors for cesarean delivery are important to practicing obstetricans. Journal of Reproductive Medicine 2006;In Press.
10. Harrell F. Regression Modeling Strategies with Applications to Linear Models, Logistic Regression, and Survival AnalysisSpringer Series in Statistics. New York: Springer-Verlag, 2001.
11. Hosmer D, Lemeshow S. Applied Logistic Regression. New York: Wiley, 2000.
12. Efron B, Tibshriani R. An Introduction to the Bootstrap. New York: Chapman and Hall, 1993.
13. S-PLUS. Insightful Software.
14. Iezzoni L. Risk-adjustment for measuring health care outcomes. Chicago, IL: Health administration press, 2003.
15. Person SD, Allison JJ, Kiefe CI, Weaver MT, Williams OD, Centor RM, et al. Nurse staffing and mortality for Medicare patients with acute myocardial infarction.[see comment]. Medical Care 2004;42:4-12.
16. Gregory KD, Korst LM. Age and racial/ethnic differences in maternal, fetal, and placental conditions in laboring patients. American Journal of Obstetrics & Gynecology 2003;188:1602-6.
17. Bacak SJ, Callaghan WM, Dietz PM, Crouse C. Pregnancy-associated hospitalizations in the United States, 1999-2000. American Journal of Obstetrics & Gynecology 2005;192:592-7.
18. Yasmeen S, Romano PS, Schembri ME, Keyzer JM, Gilbert WM. Accuracy of obstetric diagnoses and procedures in hospital discharge data. American Journal of Obstetrics & Gynecology 2006;194:992-1001.
19. Yeast JD, Jones A, Poskin M. Induction of labor and the relationship to cesarean delivery: A review of 7001 consecutive inductions.[see comment]. American Journal of Obstetrics & Gynecology 1999;180:628-33.