Risk Adjusting Cesarean Delivery Rates: A Comparison of Hospital Profiles Based on Medical Record and Birth Certificate Data

David L. DiGiuseppe, David C. Aron, Susan M. C. Payne, Richard J. Snow, LeRoy Dierker, and Gary E. Rosenthal

Objectives. Compare the discrimination of risk-adjustment models for primary cesarean delivery derived from medical record data and birth certificate data and determine if the two types of models yield similar hospital profiles of risk-adjusted cesarean delivery rates.

Data Sources/Study Setting. The study involved 29,234 women without prior cesarean delivery admitted for labor and delivery in 1993–95 to 20 hospitals in northeast Ohio for whom data abstracted from patient medical records and data from birth certificates could be linked.

Study Design. Three pairs of multivariate models of the risk of cesarean delivery were developed using (1) the full complement of variables in medical records or birth certificates; (2) variables that were common to the two sources; and (3) variables for which agreement between the two data sources was high. Using each of the six models, predicted rates of cesarean delivery were determined for each hospital. Hospitals were classified as outliers if observed and predicted rates of cesarean delivery differed ($p < .05$).

Principal Findings. Discrimination of the full medical record and birth certificate models was higher ($p < .001$) than the discrimination of the more limited common and reliable variable models. Based on the full medical record model, six hospitals were classified as statistical ($p < .01$) outliers (three high and three low). In contrast, the full birth certificate model identified five low and four high outliers, and classifications differed for seven of the 20 hospitals. Even so, the correlation between adjusted hospital rates was substantial ($r = .71$). Interestingly, correlations between the full medical record model and the more limited common ($r = .84$) and reliable ($r = .88$) variable birth certificate models were higher, and differences in classification of hospital outlier status were fewer.

Conclusion. Birth certificates can be used to develop cesarean delivery risk-adjustment models that have excellent discrimination. However, using the full complement of birth certificate variables may lead to biased hospital comparisons. In contrast, limiting models to data elements with known reliability may yield rankings that are more similar to rankings based on medical record data.

Key Words. Birth certificates, cesarean section, provider profiling, quality of care, risk adjustment

---

The cesarean section (C-section) delivery rate is increasingly being used as a measure of hospital and health plan performance (New England HEDIS Coalition 1994; Gabay and Wolfe 1994) with the implicit assumption that lower rates indicate more efficient care, more appropriate care, or both. Although unadjusted C-section rates are often used in such profiles, two recent analyses demonstrated poor agreement between unadjusted hospital rates and rates that were adjusted for maternal and neonatal risk factors that may increase the likelihood or necessity of a cesarean delivery (Aron et al. 1998; Bailit, Dooley, and Peaceman 1999). Both of these studies attempted to decrease the bias introduced by differences in patient mix. However, the first used data elements that were abstracted from patients' medical records, and the second used data from birth certificates.

Although the use of medical record data has been considered preferable in risk adjusting patient outcomes (Iezzoni 1994), the cost of abstracting medical records is substantial and may be prohibitive for large-scale initiatives to examine cesarean delivery rates. Alternatively, birth certificate data are readily available in all states and can be obtained at low cost. Such databases blend the characteristics of an administrative database with those of a clinical database and include many potential maternal and neonatal physiologic risk

---

This work was supported in part by a Career Development Award to Dr. Rosenthal from the Health Services Research and Development Service, U. S. Department of Veterans Affairs and a grant from the Medicaid Technical Assistance and Policy Program, Ohio Department of Human Services.

David L. DiGiuseppe, M.S. is an independent consultant, Seattle, WA. David C. Aron, M.D. is Professor, Division of Clinical and Molecular Endocrinology, Department of Medicine, Case Western Reserve University, and Cleveland VA Medical Center, Cleveland, OH. Susan M. C. Payne, Ph.D., M.P.H. is Associate Professor, Edmund S. Muskie School of Public Service, University of Southern Maine, Portland. Richard J. Snow, D.O. is Adjunct Assistant Professor, Department of Epidemiology and Biostatistics, Case Western Reserve University. LeRoy Dierker, M.D., L.D. is Professor, Department of Reproductive Biology, Case Western Reserve University, and Director, Division of Obstetrics, MetroHealth Medical Center, Cleveland, OH. Address correspondence to Gary E. Rosenthal, M.D., Associate Professor and Director, Division of General Internal Medicine, University of Iowa Hospitals and Clinics, 200 Hawkins Drive, SE 618 GH, Iowa City, IA 52246. This article, submitted to *Health Services Research* on September 8, 1999, was revised and accepted for publication on May 23, 2000.

factors. Data elements may be abstracted from the medical record, obtained from patients, or directly provided by obstetric personnel involved in caring for patients (Starr and Starr 1995). The primary drawback to using birth certificate data is that data collection procedures are not standardized across hospitals. Indeed, recent studies suggest that the reliability of birth certificate data may be poor for certain elements, especially maternal comorbid conditions and complications of pregnancy (e.g., abruptio placentae, umbilical cord prolapse) (Piper, Mitchel, Snowden, et al. 1993; Parrish, Holt, Connell, et al. 1993; Buescher et al. 1993)—factors that may be particularly useful in risk-adjustment models for cesarean delivery. To our knowledge no previous studies have directly compared agreement between cesarean delivery rates that were adjusted using birth certificate data and rates adjusted using medical record data.

We conducted the current study to compare the discrimination of risk-adjustment models for cesarean delivery that were developed using data abstracted from the medical record to models developed using birth certificate data for the same deliveries and to determine if hospitals identified as statistical outliers varied according to the data source used to develop the model.

METHODS

Hospitals

The study was conducted in 20 hospitals in the Cleveland metropolitan area that provided obstetric services and participated in Cleveland Health Quality Choice (CHQC), a regional coalition of employers, hospitals, and physicians to compare hospital performance (Rosenthal and Harper 1994). The total number of deliveries performed per hospital during the three-year study period (January 1993 through December 1995) ranged from 478 to 12,970 (median 3,424). Six of the 20 hospitals in the sample were classified as teaching, including five hospitals with accredited residency training programs in obstetrics and gynecology and one hospital in which residents from other clinical services rotated on the obstetric service.

Patients

The CHQC sample was drawn from consecutive eligible women admitted for labor and delivery to the study hospitals during the study period ($N =$ 89,676). Patients were identified on the basis of specific ICD-9-CM principal diagnosis codes 650 to 674. Patients undergoing therapeutic abortions or who did not deliver at least one infant weighing 500 g or more were ineligible for data collection.

CHQC implemented a protocol for sampling eligible patients in each of the participating hospitals that was dependent on obstetric volume. The protocol was implemented to decrease the costs of data collection in the higher-volume hospitals. In hospitals performing fewer than 1,200 deliveries per year (n = 10), complete study data were abstracted from the medical records of all eligible patients (n = 17,739). In hospitals performing more than 1,200 deliveries per year (n = 10), complete data were collected on a random sample of roughly 600 patients per hospital per year (n = 18,035) as previously described (Aron et al. 1998). The random sample represented 26 percent of all deliveries performed in those hospitals over the study period. In all 20 hospitals complete CHQC data were available for 35,774 patients.

Data Collection

CHQC data were abstracted on standard forms by trained medical record technicians at each hospital. Data elements included maternal demographics, pre-existing comorbid illnesses (e.g., diabetes mellitus), pregnancy-induced conditions (e.g., hypertension), obstetric conditions (e.g., abruptio placentae, cord prolapse), maternal and neonatal clinical findings (e.g., maternal blood pressure and hematocrit, fetal heart rate and birth weight), and maternal and neonatal outcomes (e.g., delivery type, infant disposition). Procedures to ensure reliability of the CHQC data have been reported previously (Rosenthal and Harper 1994; Aron et al. 1998).

Birth certificate data for the study hospitals were obtained from the Ohio Department of Health (ODH) as electronic files. Variables included maternal demographics, prior obstetric history, substance abuse, receipt of prenatal care, pre-existing and pregnancy-related illnesses, obstetric conditions, gestational age, type of delivery, birth weight, Apgar score, complications of labor and delivery, congenital anomalies, and maternal and neonatal disposition.

Analytic Cohort

Neither the CHQC nor ODH database included unique patient identifiers. The two sources were merged using several common variables including the hospital in which the delivery occurred, maternal birth date, delivery date, and infant birth weight. Of the 35,774 CHQC records for which complete data existed, 35,245 (98.5 percent) were successfully matched to a unique ODH birth certificate record. The 529 records that were not successfully matched were excluded from the analysis. Because the risk-adjustment models were developed to predict primary cesarean delivery, we excluded 5,654 deliveries in which either the CHQC or the ODH data indicated that the patient had had a prior C-section. We further excluded 357 deliveries in which either data source indicated a multiple gestation, leaving a final analytic cohort of 29,234.

Risk-Adjustment Models

We developed separate risk-adjustment models from the CHQC medical record and ODH birth certificate data. The "full" medical record model was derived from 38 of the 39 risk factors used in our previous analysis (see Table 1) (Aron et al. 1998). Multiple gestation was not included in the model because such patients were excluded from the current cohort. Each model variable was associated ($p < .1$) with cesarean delivery in bivariate analyses using the chi-square test. These variables were then entered into a logistic regression model from which a predicted risk of cesarean delivery was determined for each patient.

The full birth certificate model was derived in a similar manner. Bivariate associations between cesarean delivery and 63 maternal and neonatal risk factors available in the birth certificate data set were examined using the chi-square test. Variables found to be associated ($p < .1$) with C-section delivery were included as independent variables in a logistic regression model, with the exception of cephalopelvic disproportion, fetal distress, anesthetic complications, and "other complications of labor and/or delivery," which were believed to be either particularly difficult to standardize across hospitals or reflect the quality of predelivery care (Table 1). The full birth certificate model included 34 risk factors as independent variables, many of which were similar to variables included in the medical record model. Two additional models were developed based on the 22 data elements that were common to the two sources—the "common variable" medical record and birth certificate models. Last, we developed a third pair of models that were restricted only to the eight data elements that were common to the two data sources and demonstrated high reliability (i.e., kappa statistic > .60)—the "reliable variable" medical record and birth certificate models.

Discrimination of the risk-adjustment models was measured using the c statistic, which represents the proportion of times patients undergoing cesarean delivery had higher predicted risks of cesarean delivery than patients undergoing vaginal delivery. Pairwise comparisons of c statistics for different

Table 1: Prevalence, Cesarean Delivery Rates, and Adjusted Odds Ratios for Risk Factors Included in Full Medical Record and Birth Certificate Models, as Determined by Multiple Logistic Regression

	Prevalence, %		*C-Section Rate (%) when Risk Factor Present*		*Adjusted Odds Ratio from Base Model*	
Risk Factor	*Medical Records*	*Birth Certificates*	*Medical Records*	*Birth Certificates*	*Medical Records*	*Birth Certificates*
Maternal age (years)						
<18	5.2	5.2	12.8	12.8	referent	referent
18–20†,‡	12.3	12.3	11.8	11.7	1.07 (0.87,1.32)	1.11 (0.90,1.36)
21–24†,‡	19.8	19.8	12.3	12.4	1.42 (1.16,1.73)	1.55 (1.28,1.89)
25–28†,‡	22.5	22.5	15.0	15.1	1.87 (1.54,2.27)	2.13 (1.76,2.58)
29–33†,‡	26.6	26.6	13.7	13.7	1.90 (1.57,2.31)	2.25 (1.86,2.73)
34–38†,‡	11.4	11.5	14.7	14.7	2.57 (2.07,3.18)	3.02 (2.44,3.74)
≥39†,‡	2.2	2.2	20.1	19.9	3.56 (2.62,4.82)	4.73 (3.50,6.39)
Maternal weight						
Prepregnancy ≥ 250 lbs.	N/A	1.4	N/A	28.2	N/A	2.39 (1.84,3.11)
Gain during pregnancy ≥ 50 lbs.	N/A	8.6	N/A	21.9	N/A	1.44 (1.27,1.63)
Pregnancy history						
Nulliparity[a],†,‡	47.8	46.1	22.2	22.7	6.71 (6.07,7.43)	6.51 (5,89,7.20)
Prior infant > 4,000 g	N/A	0.9	N/A	6.7	N/A	0.66 (0.38,1.16)
Prior infant small for gestational age	N/A	1.4	N/A	6.0	N/A	0.61 (0.36,1.03)
Pre-existing comorbid conditions						
Diabetes mellitus	0.4	N/A	37.6	N/A	2.49 (1.53,4.03)	N/A
Hypertension†	0.7	0.6	28.3	27.4	1.86 (1.27,2.72)	1.46 (0.96,2.22)
Severe comorbid illness[b]	0.5	N/A	23.5	N/A	1.67 (1.07,2.60)	N/A
Anemia	4.1	N/A	16.6	N/A	1.42 (1.18,1.72)	N/A
Thyroid disease	1.0	N/A	19.5	N/A	1.40 (0.97,2.00)	N/A
Lung disease†	3.1	1.2	15.8	17.2	0.99 (0.79,1.24)	1.11 (0.77,1.58)

Tobacco use[†,‡]	24.1	18.5	12.1	12.0	0.94 (0.85,1.04)	1.01 (0.91,1.13)
Alcohol use	N/A	2.2	N/A	11.0	N/A	0.73 (0.53,1.00)
Hemoglobinopathy	N/A	0.0	N/A	13.7	N/A	2.47 (0.55,11.17)
Renal disesae	N/A	0.3	N/A	20.4	N/A	1.59 (0.87,2.91)
Other medical risk factors this pregnancy[c]	N/A	14.5	N/A	17.8	N/A	1.47 (1.33,1.64)
Conditions arising during pregnancy						
Genital herpes[†]	1.2	0.7	32.2	33.0	3.79 (2.91,4.93)	3.30 (2.34,4.73)
Other sexually transmitted diseases[d]	3.5	N/A	10.8	N/A	0.83 (0.65,1.05)	N/A
Gestational diabetes[†]	3.7	2.4	23.2	26.1	1.72 (1.44,2.05)	1.84 (1.49,2.27)
Pregnancy-induced hypertension	4.0	2.4	30.0	33.2	1.66 (1.39,1.97)	2.27 (1.87,2.75)
Gestational age						
Term (37–41 weeks)	90.6	90.3	12.7	12.9	referent	referent
Postterm (≥42 weeks)[†,‡]	3.3	3.8	30.7	24.4	3.05 (2.58,3.59)	2.01 (1.70,2.37)
Preterm (<37 weeks)[†,‡]	6.1	5.9	18.6	18.9	0.92 (0.75,1.13)	1.17 (0.99,1.39)
Infant body weight > 4,000 g[†,‡]	10.9	11.0	21.2	21.4	2.25 (2.01,2.52)	2.13 (1.89,2.38)
Obstetrical risk factors and conditions						
Breech presentation[e,†,‡]	3.0	3.3	88.5	88.8	120.09 (95.20,151.48)	111.08 (88.69,139.14)
Face or transverse presentation[†,‡]	0.6	(e)	62.4	(e)	21.19 (14.50,30.95)	(e)
Placenta previa[†]	0.5	0.3	56.0	81.1	15.33 (10.43,22.54)	60.45 (33.51,109.06)
Umbilical cord prolapse[†]	0.2	0.1	76.1	72.1	26.62 (14.08,50.34)	19.96 (9.25,43.07)
Abruptio placentae[†]	0.7	0.6	49.8	55.8	8.39 (6.03,11.68)	15.10 (10.51,21.69)
Eclampsia/pre-eclampsia[†]	2.2	0.2	39.7	44.8	2.97 (2.41,3.66)	3.30 (1.77,6.16)
Third-trimester bleeding[f,†]	1.3	1.1	29.1	21.6	2.19 (1.64,2.93)	0.88 (0.61,1.26)
Oligohydramnios[g,†]	1.7	0.7	36.2	36.6	2.74 (2.16,3.48)	3.22 (2.24,4.62)
Polyhydramnios[†]	0.4	(g)	29.0	(g)	2.60 (1.61,4.21)	(g)
Chorioamnionitis	5.4	N/A	29.2	N/A	2.42 (2.12,2.78)	N/A
Intrauterine growth retardation	1.0	N/A	26.1	N/A	1.37 (0.98,1.93)	N/A
Premature rupture of membranes[†]	2.1	3.2	26.1	23.6	2.15 (1.73,2.67)	1.31 (1.08,1.58)

Continued

Table 1: *Continued*

Risk Factor	Prevalence, %		C-Section Rate (%) when Risk Factor Present		Adjusted Odds Ratio from Base Model	
	Medical Records	Birth Certificates	Medical Records	Birth Certificates	Medical Records	Birth Certificates
Meconium staining/aspiration[†]	14.0	7.3	20.1	19.8	1.64 (1.48,1.81)	1.57 (1.37,1.79)
Cervical incompetence	0.5	N/A	20.9	N/A	1.64 (0.97,2.78)	N/A
Seizures	N/A	0.0	N/A	66.7	N/A	11.45 (1.40,93.86)
Risk factors related to labor						
Premature labor	6.7	N/A	13.4	N/A	0.68 (0.55,0.84)	N/A
Precipitous labor	N/A	2.1	N/A	0.5	N/A	0.04 (0.01,0.15)
Prolonged labor	N/A	0.8	N/A	45.1	N/A	3.00 (2.22,4.05)
Dysfunctional labor	N/A	3.0	N/A	42.3	N/A	4.21 (3.58,4.94)
Fetal abnormalities						
Fetal anomalies diagnosed prior to birth[h,†]	0.3	N/A	31.2	N/A	3.10 (1.72,5.57)	N/A
Hydrocephalus[†]	N/A	0.0	N/A	66.7	N/A	12.56 (1.96,80.55)
Diaphragmatic hernia[†]	N/A	0.0	N/A	37.5	N/A	2.15 (0.49,9.50)
Other gastrointestinal abnormality[†]	N/A	0.0	N/A	37.5	N/A	3.20 (0.34,29.91)
Cleft lip/palate[†]	N/A	0.1	N/A	29.2	N/A	1.56 (0.43,5.62)
Admission vital signs						
Maternal systolic blood pressure						
<140 mm Hg	84.1	N/A	12.8	N/A	referent	N/A
140–159 mm Hg	14.4	N/A	17.6	N/A	1.09 (0.98,1.22)	N/A
≥160 mm Hg	1.5	N/A	29.1	N/A	1.65 (1.26,2.15)	N/A
Maternal heart rate ≥100 bpm	17.8	N/A	15.3	N/A	1.17 (1.05,1.30)	N/A

Fetal heart rate						
120–159 bpm	94.5	N/A	13.5	N/A	referent	N/A
≥ 160 bpm	3.4	N/A	16.9	N/A	1.29 (1.05,1.59)	N/A
<120 bpm	2.1	N/A	17.9	N/A	1.66 (1.30,2.11)	N/A
Maternal fever	N/A	1.6	N/A	29.2	N/A	1.66 (1.32,2.09)
Maternal hemoglobin < 90 g/L	0.8	N/A	12.5	N/A	0.93 (0.57,1.54)	N/A

[a]Birth certificate variable determined from the following variables: "date of last live birth," "live births now living," and "live births now dead."

[b]Includes the following conditions: acquired immunodeficiency syndrome, coagulation disorder, collagen vascular disease, malignancy, sickle cell disease, and thalassemia.

[c]Represents presence of other unspecified risk factors that are not explicitly listed on the birth certificate worksheet.

[d]Includes gonorrhea, syphilis, and chlamydial infections.

[e]Birth certificate records "breech/malpresentation" as a single variable.

[f]Birth certificate variable determined from the variable, "uterine bleeding."

[g]Birth certificate records "polyhydramnios and oligohydramnios" as a single variable.

[h]Includes the following conditions: cleft palate, congenital heart disease, congenital dislocation of hip, Down's syndrome, hydrocephalus, spina bifida, talipes equinovarus, pyloric stenosis, or other abnormalites; birth certificate analog determined from list of "congenital anomalies of child" including: anencephalus, spina bifida/meningocele, hydrocephalus, microcephalus, other central nervous system anomalies, heart malformations, other circulatory/respiratory anomalies, rectal atresia/stenosis, tracheo-esophageal fistula/esophageal atresia, omphalocele/gastroschisis, other gastrointestinal anomalies, malformed genitalia, renal agenesis, other urogenital anomalies, cleft lip/palate, polydactyly/syndactyly/adactyly, club foot, diaphragmatic hernia, other musculoskeletal/integumental anomalies, Down's syndrome, and other chromosomal anomalies.

†Variable also included in common variable medical record and birth certificate models.

‡Variable also included in reliable variable medical record and birth certificate models.

models were made using the method of Hanley and McNeil (1983) modified for correlated data.

Our objective was to explain as much of the variation in cesarean delivery rates in the current cohort as possible prior to determining hospital-level effects, and not necessarily to export the model to other populations. We were therefore less concerned with model overfitting than with maximizing discrimination. Nevertheless, we determined cross-validated c statistics for each model to examine the stability of model predictions by randomly splitting the analytic cohort into equal-sized ($n = 14{,}617$) "development" and "validation" samples. Using the variables from the full-sample model, a new model was fit to the development sample only, and the coefficients from this model were applied to patients in the validation sample. The resulting predicted risks of cesarean delivery were then used to generate a c statistic. The roles of the development and validation samples were reversed and the above procedures repeated. The mean of the two c statistics represented the cross-validated c statistic.

Analysis

Predicted risks of cesarean delivery in individual patients were aggregated by hospital to determine the mean predicted risk for each hospital for each of the six risk-adjustment models. For each hospital, the observed cesarean delivery rate was compared to predicted cesarean delivery rates based on each of the six risk-adjustment models by calculating a z statistic based on the following equation:

$$Z_h = \frac{(O_h - P_h)}{\sqrt{\frac{\sum [P_j * (1 - P_j)]}{N_h}}}$$

where O_h represents the hospital's observed rate, P_h represents the hospital's mean predicted rate, P_j represents the predicted score for the jth delivery at the hospital, and N_h represents the hospital's total number of deliveries (Iezzoni et al. 1996). For each model hospitals were classified as statistical outliers if the z score was greater than 1.96 or less than -1.96.

Determining outlier status based on arbitrary cut points such as $p < .05$ may be misleading as small differences in z scores may lead to movement past outlier thresholds. To further examine differences between the risk-adjustment models we determined Pearson's product-moment correlations between hospital z scores based on different models and Spearman's rankorder correlations between z score rankings (i.e., 1 to 20). All statistical analyses were conducted using SAS for Windows, Version 6.12.

RESULTS

The mean age of the 29,234 patients in the study sample was 26.7 ± 5.9 years. Based on information in the CHQC medical record database, 75.5 percent of patients were Caucasian and 19.2 percent were African American; in 5.3 percent of patients, race was classified as other or unknown. In addition, 65.9 percent of patients had commercial insurance (including indemnity and managed care plans), 27.6 percent had Medicaid or another governmental form of insurance (Medicare, county assistance, workers compensation), and 4.3 percent were uninsured; for 2.1 percent of patients, insurance status was classified as other or unknown. Nearly all (97 percent) received some prenatal care prior to delivery. For the 27,001 patients for whom prenatal care was provided and the trimester of first occurrence was known, 71.7 percent had received care during the first trimester. Histories of chronic tobacco, alcohol, or illegal drug use were documented in 24.1 percent, 7.1 percent, and 2.6 percent of patients, respectively. The overall primary C-section rate was 13.7 percent ($n = 4,005$) and ranged from 5.7 to 18.9 percent in individual hospitals; five hospitals had rates of less than 12 percent, whereas four hospitals had rates greater than 16 percent.

The prevalence of individual risk factors in the medical record and birth certificate databases and rates of cesarean delivery in deliveries in which the risk factor was present are shown in Table 1. The prevalence of risk factors that were common to the two data sets was generally similar. Relative differences were most notable for a lower prevalence of the following variables in birth certificate data: lung disease (1.2 percent vs. 3.1 percent in medical record data), genital herpes (0.7 percent vs. 1.2 percent), pregnancy-induced hypertension (2.4 percent vs. 4.0 percent), and meconium staining or aspiration (7.3 percent vs. 14.0 percent).

Rates of cesarean delivery were generally higher in deliveries in which the factors noted in Table 1 were present. The highest rates were seen in deliveries with breech presentation (88.5 percent per medical record and 88.8 percent per birth certificate), seizures during labor (66.7 percent per birth certificate), hydrocephalus (66.7 percent per birth certificate), face or transverse presentation (62.4 percent per medical record), placenta previa (56.0 percent and 81.1 percent, respectively), cord prolapse (76.1 percent and 72.1 percent, respectively), and abruptio placenta (49.8 percent and 55.8 percent, respectively).

Using logistic regression analysis the six risk-adjustment models were estimated (Table 1). Three variables yielded multivariate odds ratios of 3.0 or greater in all six models: breech presentation (estimated odds ratios ranged from 90.0 to 120.1), nulliparity (6.5 to 7.3), and age > 39 years (3.6 to 4.7). Risk factors with odds ratios above 3.0 in the four full and common risk-adjustment models included placenta previa (14.6 to 65.6), cord prolapse (20.0 to 28.0), abruptio placentae (8.4 to 15.1), genital herpes (3.3 to 3.8), and congenital anomalies (3.1 to 12.6). Discrimination of all six models was excellent, with c statistics of .849 for the full medical record model, .846 for the full birth certificate model, .841 for the common variable medical record model, .829 for the common variable birth certificate model, .801 for the reliable variable medical record model, and .800 for the reliable variable birth certificate model. Cross-validation c statistic values were nearly identical to the overall c statistics (.847, .843, .840, .827, .801, and .800, respectively), indicating the stability of the model variables in the study population. Because of the high degree of correlation between the six models, and despite the relatively narrow range in c statistics, discrimination of the four common and reliable variable models was lower ($p < .001$) than for the full medical record model. Evaluating the models as pairs, discrimination of the full medical record and full birth certificate models was similar ($p = .39$), as was the discrimination of the reliable variable medical record and reliable variable birth certificate models ($p = .73$), although discrimination of the medical record common variable model was higher ($p < .001$) than the discrimination of the birth certificate common variable model.

The identification of hospitals as statistical outliers ($p < .05$) differed depending on the specific risk-adjustment model that was applied. The full medical record model identified three low and three high outliers. In contrast, the full birth certificate model identified five low and four high outliers.

Seven of the 20 hospitals were classified differently based on the two models. Two hospitals classified as a low ($n = 1$) or high ($n = 1$) outlier based on the medical record model but were nonoutliers based on the birth certificate model; five hospitals were nonoutliers based on the medical record model but were low ($n = 3$) or high ($n = 2$) outliers based on the birth certificate model.

The common variable medical record model identified three low and four high outliers, whereas the common variable birth certificate model identified five low and five high outliers. Classification differed for only three hospitals, including two low outlier hospitals and one high outlier hospital based on the birth certificate model that were nonoutliers based on the medical record model.

The reliable variable medical record model identified five low and three high outliers, whereas the birth certificate model identified five low and four high outliers. Classification differed for one hospital, a high outlier based on birth certificates that was a nonoutlier based on medical records.

Classifications based on the full medical record model and on the common variable and reliable variable birth certificate models were also compared. Outlier classifications differed for four hospitals based on the common variable model and for only three hospitals based on the reliable variable model. Interestingly, differences with the full medical record model were less with these more limited birth certificate models than with the more complete full birth certificate model.

Correlations between hospital z scores for the six models were substantial, ranging from .71 to .98 (see Table 2). Correlations between hospital rankings were more moderate (.55 to .97) although patterns were similar. As expected, correlations were highest between models based on similar data sources. Correlations were lowest between the full birth certificate model and the three medical record models. Similar to the results of analyses comparing outlier classifications, correlations between z scores for the full medical record model and the common and reliable variable birth certificate models were higher than the correlation between z scores for the full medical record and birth certificate models. Finally, differences between hospital z scores for the full medical record and the three birth certificate models tended to be higher for the full birth certificate model (mean difference 1.82 ± 1.66) than for the common variable (mean difference 1.46 ± 1.24) or reliable variable (mean difference 1.28 ± 1.01) birth certificate models.

DISCUSSION

The current study represents the first comparative evaluation of alternative methods for risk adjusting hospital primary cesarean delivery rates. Analyzing patients admitted to 20 hospitals in a single metropolitan area over a three-year period, we found that risk-adjustment models developed from readily available birth certificate data had similar discrimination to models based on data abstracted from patient medical records. In addition, correlations between hospital z scores and rankings based on the different models were

Table 2: Correlation Between Hospital z Scores and Hospital Rankings (i.e., 1–20) Based on the Six Multivariable Risk-Adjustment Models for Primary Cesarean Delivery ($p < .01$ for All Values)

	Pearson/Spearman Correlation Coefficient				
	Full Birth Certificate	*Common Variable Medical Record*	*Common Variable Birth Certificate*	*Reliable Variable Medical Record*	*Reliable Variable Birth Certificate*
Full medical records	.71/.55	.98/.97	.84/.66	.94/.88	.88/.80
Full birth certificate	—	.71/.55	.95/.93	.79/.63	.82/.68
Common variable medical record		—	.83/.65	.94/.86	.86/.76
Common variable birth certificate			—	.92/.73	.95/.79
Reliable variable medical record				—	.97/.96

moderate to substantial. However, the classification of hospitals as statistical outliers differed markedly depending on the particular risk-adjustment method used. The differences were most pronounced for models based on medical record data and a set of variables based on the full spectrum of potential risk factors available in birth certificates. Differences were less pronounced between medical record and birth certificate models that were based only on variables common to both data sources or on variables in which agreement between the two sources was high. The more limited birth certificate models yielded classifications of outlier status that were more similar to the full medical record model than the full birth certificate model.

If information abstracted from patient medical records represents a gold-standard method for risk adjustment, our findings indicate that many of the variables available in birth certificate data may not be suitable to use in risk adjustment. However, using a more limited set of birth certificate variables, particularly those for which reliability is high, may represent a reasonable alternative to medical record data for comparing hospital-level primary cesarean delivery rates.

In addition to birth certificate data, other existing data sources may be available for risk adjusting hospital cesarean delivery rates. For example, hospital discharge abstract (i.e., claims) data include ICD-9-CM codes for many of the factors included in our medical record and birth certificate risk-adjustment models (e.g., placental abnormalities, pregnancy-induced hypertension, gestational diabetes). However, the use of claims data in evaluating other outcomes (e.g., hospital mortality and costs) has been met by substantial criticism from clinicians and hospital administrators, and a large body of literature has demonstrated the theoretical and practical limitations of using claims data in risk adjustment (Berwick and Wald 1990; Iezzoni 1994). Nevertheless, given the increasing availability of discharge abstract databases on state or regional levels, examining the validity and reliability of claims data for obstetric patients and determining whether discharge abstract data may improve risk adjustment based on birth certificates would represent promising areas of investigation.

The differences we found between the risk-adjustment models in classifying hospitals as outliers in this study may be because of several factors. Although the medical record and birth certificate models included many of the same variables, prior studies indicate that agreement between birth certificate and medical record data is often poor, particularly for maternal comorbid conditions (e.g., hypertension, anemia) and obstetric risk factors (e.g., placenta previa, umbilical cord prolapse) (Piper, Mitchel, Snowden, et al. 1993; Parrish, Holt, Connell, et al. 1993;, Buescher et al. 1993). This may reflect the lack of standardization in the collection and recording of birth certificate variables (Parrish, Holt, Connell, et al. 1993). To further examine this possibility we compared medical record and birth certificate models that only included age and 21 other risk factors that were common to both data sources. Even these reduced models led to outlier classifications that differed for three of the 20 hospitals, indicating discrepancies between the medical records and birth certificates in how these variables are recorded. As would be expected, further limiting the models only to data elements for which there was substantial agreement between the two data sources yielded hospital classifications that were nearly identical.

Our finding that the classification of hospitals as outliers varied for different risk-adjustment models parallels the results of studies of hospital mortality. In analyses of pneumonia, coronary artery bypass graft surgery, acute myocardial infarction, and stroke, Iezzoni and colleagues have reported that different severity measures applied to the same group of patients lead to identification of different hospitals as statistical outliers (Iezzoni, Ash, Shwartz, et al. 1995; Iezzoni et al. 1996; Iezzoni, Ash, Shwartz, et al. 1996; Iezzoni, Shwartz, Ash, et al. 1996; Landon, Iezzoni, Ash, et al. 1996). For example, in a study of patients hospitalized for pneumonia (Iezzoni et al. 1996), risk-adjusted death rates differed for nearly 30 percent of hospitals depending on which of 14 different methods was used.

Our finding that risk-adjustment methods based on the unrestricted use of secondary data may yield models with similar or higher discrimination than models based on medical record data or on a restricted set of secondary data is also consistent with prior studies. In an analysis of patients with acute myocardial infarction, Iezzoni, Ash, Shwartz, et al. (1995) found that models developed from administrative data have similar or higher discrimination than models developed from medical record data. However, in an attempt to examine the attributional validity of the methods, the medical record–based methods more frequently yielded a higher probability of mortality when six clinical findings believed to represent severe illness in acute myocardial infarction were present. In a second study, Pine et al. (1997) found that unrestricted administrative risk-adjustment models (average *c* statistic .87) have higher discrimination than restricted models (average *c* statistic .75) and similar discrimination as medical record models (average *c* statistic .87). However, the high discrimination of the unrestricted administrative models stemmed from the inclusion of hospital-acquired complications, and hospital rankings based on restricted administrative models yield hospital rankings more similar to rankings based on medical record data.

In interpreting our findings it is important to consider that our study was based on birth certificate data from a single metropolitan area in a single state. Although the hospitals in the study were diverse in size and other characteristics, the quality of the collection of birth certificate data in our hospitals may not be representative of birth certificate data in other regions or states. In addition, the medical record data used in our analysis may not represent a gold standard by which to compare birth certificate data or risk-adjusted cesarean delivery rates. Although explicit protocols were developed to ensure the reliability of the CHQC medical records abstraction process, the records themselves may be subject to inclusion of errant information or exclusion of important findings.

It is also important to recognize that the costs of implementing a comprehensive medical record abstraction process as in the current study are substantial. Moreover, obtaining cooperation from competing hospitals in a common health care market to implement a standardized process of abstracting medical records may be difficult, particularly in the absence of organized support from health care purchasers or physicians (Iglehart 1988; Shaller and Woods 1991; Berwick 1991; Burton 1999). In contrast, birth certificate information is readily available in all states. Our analyses indicate that careful selection of birth certificate variables may yield risk-adjustment models for cesarean delivery that provide relatively similar hospital rankings to models based on medical record data. In contrast, models based on all birth certificate variables may yield hospital rankings that exhibit more variation.

Our findings also demonstrate that solely relying on predictive validity (i.e., c statistic values) of a risk-adjustment methodology may be unwise. Although we were able to develop a birth certificate model that had similar discrimination as a model based on the full complement of data in patient medical records, a birth certificate model that used fewer, "more reliable" variables yielded rankings closer to those based on the medical record model. Although discrimination of the simpler birth certificate model was lower than the full birth certificate model, predictions based on the simpler model had greater validity when compared to a benchmark model based on the full spectrum of medical records data. Thus, when selecting risk-adjustment methods, a comprehensive approach that includes consideration of the content and face validity of alternative methods and compares the provider profiles resulting from these approaches is warranted.

Last, as reporting of risk-adjusted outcomes becomes mandated by the Joint Commission on the Accreditation of Healthcare Organizations and other organizations (Zeglen 1997; National Committee for Quality Assurance 1993), our findings indicate that methods based on different data sources or risk factors may provide different views of performance. Given that mandates may allow providers to select from multiple "certified" methods, providers may be able to choose methods that yield the most optimal results.

Risk adjustment is an important component in developing unbiased comparisons of cesarean delivery rates. Although birth certificate data are readily available in electronic formats, the development of valid hospital comparisons may require investments to abstract risk-factor data from medical records, improve the existing infrastructure for collecting birth certificate data, or identify birth certificate variables that are reliable. Such investments may in the long run be less expensive for hospitals and purchasers than making incorrect decisions on the basis of biased models.

REFERENCES

Aron, D. C., D. L. Harper, L. B. Shepardson, and G. E. Rosenthal. 1998. "Impact of Risk-Adjusting Cesarean Delivery Rates when Reporting Hospital Performance." *Journal of the American Medical Association* 279 (24): 1968–72.

Bailit, J. L., S. L. Dooley, and A. N. Peaceman. 1999. "Risk Adjustment for Interhospital Comparison of Primary Cesarean Rates." *Obstetrics and Gynecology* 93 (6): 1025–30.

Berwick, D. M. 1991. "The Double Edge of Knowledge." *Journal of the American Medical Association* 266 (6): 841–42.

Berwick, D. M., and D. L. Wald. 1990. "Hospital Leaders' Opinions of the HCFA Mortality Data." *Journal of the American Medical Association* 263 (2): 247–49.

Buescher, P. A., K. P. Taylor, M. H. Davis, and J. M. Bowling. 1993. "The Quality of the New Birth Certificate Data: A Validation Study in North Carolina." *American Journal of Public Health* 3 (8): 1163–65.

Burton, T. M. 1999. "Examining Table: Operation that Rated Hospitals Was Success, but the Patience Died." *The Wall Street Journal*, August 23, p. 1.

Gabay, M., and S. M. Wolfe. 1994. *Unnecessary Cesarean Sections: Curing a National Epidemic.* Washington, DC: Public Citizen Health Research Group.

Hanley, J. A., and B. J. McNeil. 1983. "A Method of Comparing the Areas Under Receiver Operating Characteristic Curves Derived from the Same Cases." *Radiology* 148 (3): 839–43.

Iezzoni, L. I. 1994. *Risk Adjustment for Measuring Health Care Outcomes.* Chicago: Health Administration Press.

Iezzoni, L. I., A. S. Ash, M. Shwartz, J. Daley, J. S. Hughes, and Y. D. Mackiernan. 1995. "Predicting Who Dies Depends on How Severity Is Measured: Implications for Evaluating Patient Outcomes." *Annals of Internal Medicine* 123 (10): 763–70.

———. 1996. "Judging Hospitals by Severity-Adjusted Mortality Rates: The Influence of the Severity-Adjustment Method." *American Journal of Public Health* 86 (10): 1379–87.

Iezzoni, L. I., M. Shwartz, A. S. Ash, J. S. Hughes, J. Daley, and Y. D. Mackiernan. 1996. "Severity Measurement Methods and Judging Hospital Death Rates for Pneumonia." *Medical Care* 34 (1): 11–28.

Iezzoni, L. I., M. Shwartz, A. S. Ash, and Y. D. Mackiernan. 1996. "Predicting In-Hospital Mortality for Stroke Patients: Results Differ Across Severity-Measurement Methods." *Medical Decision Making* 16 (4): 348–56.

Iglehart, J. K. 1988. "Competition and the Pursuit of Quality: A Conversation with Walter McClure." *Health Affairs* 7 (1): 79–80.

Landon, B., L. I. Iezzoni, A. S. Ash, M. Shwartz, J. Daley, J. S. Hughes, and Y. D. Mackiernan. 1996. "Judging Hospitals by Severity-Adjusted Mortality Rates: The Case of CABG Surgery." *Inquiry* 33 (2): 155–66.

National Committee for Quality Assurance. 1993. *Health Plan Employer Data and Information Set and User's Manual, Version 2.0.* Washington, DC: NCQA.

New England HEDIS Coalition. 1994. *New England HEDIS Coalition 1993 Baseline Performance Profile.* Boston: New England HEDIS Coalition.

Parrish, K. M., V. L. Holt, F. A. Connell, B. Williams, and J. P. LoGerfo. 1993. "Variations in the Accuracy of Obstetric Procedures and Diagnoses on Birth Records in Washington State, 1989." *American Journal of Epidemiology* 138 (2):119–27.

Pine, M., M. Norusis, B. Jones, and G. E. Rosenthal. 1997. "Predictions of Hospital Mortality Rates: A Comparison of Data Sources." *Annals of Internal Medicine* 126 (5): 347–54.

Piper, J. M., E. F. Mitchel, Jr., M. Snowden, H. Hall, M. Adams, and P. Taylor. 1993. "Validation of 1989 Tennessee Birth Certificates Using Maternal and Newborn Hospital Records." *American Journal of Epidemiology* 137 (7): 758–68.

Rosenthal, G. E., and D. L. Harper. 1994. "Cleveland Health Quality Choice: A Model for Community-Based Outcomes Assessment." *Joint Commission Journal on Quality Improvement* 20 (8): 425–44.

Shaller, D. V., and P. Woods. 1991. "Reforming the Market for Value in Health Care: The Cleveland Experience." *Journal of Occupational Medicine* 33 (3): 358–62.

Starr, P., and S. Starr. 1995. "Reinventing Vital Statistics: The Impact of Changes in Information Technology, Welfare Policy, and Health Care." *Public Health Reports* 110: 534–44.

Zeglen, M. 1997. "Accreditation Requirements for ORYX: The Next Evolution in Accreditation." *Journal of the American Health Information Management Association* 68 (6): 20–31.