OSHPD Postpartum Maternal Outcomes Validation Study

Core Staff

Patrick S. Romano, MD MPH; Principal Investigator
Julie A. Rainwater, PhD; Project Director
Michael E. Schembri; Statistician/Analyst

Obstetric Consultants/Advisors

Shagufta Yasmeen, MD MRCOG
William M. Gilbert, MD
Nina Boe, MD
Nancy Field, MD

Medical Record Coders

Ginger Cox, RHIT CCS; Lead Coder and Coding Educator
Della Fletcher
Ann Rivera
Carol Perry
Michele Matalone
Angie Chen
Shelley Burson

Clinical Abstractors

Janet M. Keyzer, RNC MPA; Nurse Research Coordinator
Jacqueline Beraldo, RN

Readmissions Abstractors

Vaida Vegelyte
Diana Deemer
Sal Acosta

Support Staff and Student Assistants

Banafsheh Sadeghi, MD; Graduate Student Researcher
Nhue Do
Mairin Rooney
Rowena Laber

Section 1: Executive Summary

OVERVIEW

The California Hospital Outcomes Project is an initiative mandated by the State of California and conducted by the Office of Statewide Health Planning and Development (OSHPD), to develop public reports comparing hospital outcomes for selected medical conditions and surgical procedures for patients treated in hospitals throughout California. Over the last decade, CHOP has reported risk-adjusted hospital mortality rates for heart attack and community-acquired pneumonia. In 2005, OSHPD is releasing its first report on obstetric care.

Delivery was selected as an important topic for public reporting because it is the most frequent single reason for hospitalization in California, and because complications of delivery are associated with substantial health care costs and impairment of function. In 2003, for example, there were 170,465 repairs of obstetric lacerations and 147,084 cesarean deliveries performed in California hospitals – more than any other surgical procedure. Although most women who require these procedures have excellent outcomes, a small minority experience complications that cause pain, weakness, impaired bonding with their new child, bowel or bladder problems, sexual dysfunction, rehospitalization, and even death.

This technical report, prepared for OSHPD, summarizes the validation of multiple potential measures of inpatient obstetric quality of care. These measures include two that have been endorsed by OSHPD for public reporting: risk-adjusted postpartum maternal readmission rates and risk-adjusted perineal laceration rates. Other measures were also evaluated in this validation study, but are not recommended for public reporting, including risk-adjusted rates of endometritis, wound infection, hemorrhage, and urinary tract infection. This validation study was designed by the UC Davis research team in collaboration with OSHPD staff and the AB 524 Technical Advisory Committee. It was designed to address a variety of concerns, specified in detail later in this report, about the validity of using hospital-reported ICD-9-CM codes in the California Patient Discharge Data Set to report publicly on hospital performance.

The original methodology for estimating and analyzing risk-adjusted postpartum maternal readmission rates was developed in 1996, using data on deliveries performed in 1992-1993. This developmental work is fully described in a report that was published by the OSHPD in December 1996 (Section 2, citation 19). We were subsequently asked by the OSHPD to validate the data and methodology used in this 1996 report. To simplify the task, we selected a subsample of the same records for this validation study. As described in detail in later sections of this report, we collected records from hospitals in 1998, recoded and abstracted them in 1999, and performed analyses in 2000-2001. Although some results from these analyses have already appeared in print elsewhere, this report compiles all relevant findings in a single document. We believe that the findings are still informative, despite their age, because there is no evidence of statewide improvement in the coding of obstetric records over the past decade. In addition, the mean postpartum length of stay and readmission rate have remained relatively stable over time (after some decrease in length of stay during the 1990s), suggesting that the clinical factors driving readmissions have also been relatively stable. This study remains the most comprehensive published analysis of the accuracy of ICD-9-CM coded inpatient obstetric data. However, if the OSHPD continues to use the same datasets in the same manner, it would be prudent to repeat this validation study in the future.

VALIDATION QUESTIONS AND KEY FINDINGS

The primary purpose of the validation study was to evaluate the impact of errors in hospitals' reporting of risk factors and complications on risk-adjusted outcomes estimated using OSHPD's administrative data. This is known as an evaluation of criterion validity, because data of uncertain accuracy are compared against data that are known or believed to be highly accurate. These latter data represent a "criterion standard," which is uniformly applied across all hospitals, even though it may not be perfect.

The validation study was also designed to provide some information about whether measurable differences in the process of care explain, in part, observed differences in risk-adjusted outcomes. This is known as an evaluation of construct validity, because it is based on the construct (or conceptual model) that better processes of care should lead to better outcomes, and hence worse outcomes should result from worse processes of care. To the extent that there are strong associations between process measures and risk-adjusted outcome measures, we become more confident that both sets of measures describe true quality of care.

The basic study design was a retrospective cohort of women who were admitted to acute care nonfederal hospitals in California for delivery of a child in 1992 or 1993. These women were followed for 6 weeks after delivery to identify postpartum readmissions, and their records were reviewed for 9 months before delivery to identify antepartum conditions requiring hospitalization. Hospitals, and cases within hospitals, were randomly sampled from the OSHPD's Patient Discharge Data Set. To ensure sufficient statistical power to answer the questions listed below, we oversampled women at hospitals with lower or higher than expected readmission rates, women who underwent cesarean delivery, and women who required readmission. After obtaining a complete photocopy of each sampled record, we collected data through detailed, independent review by a coding professional and a research nurse, with physician back-up as needed.

The remainder of this Section outlines the specific research questions addressed by this study, and summarizes the key findings.

Question 1. What percentage of obstetric discharges reported to OSHPD are fundamentally miscoded, in terms of whether a delivery occurred or whether the delivery type was vaginal or cesarean?

The OSHPD identifies deliveries using ICD-9-CM diagnosis codes for "complications of pregnancy, childbirth, and the puerperium." For most codes in this chapter of ICD-9-CM, the fifth digit indicates whether a patient was "delivered" or was admitted simply for an "antepartum condition or complication" or a "postpartum condition or complication." Among these deliveries, cesarean deliveries are identified by the presence of specific procedure codes, and vaginal deliveries are defined by default as all deliveries without any of those procedure codes. We evaluated the accuracy of both delivery diagnosis codes and cesarean procedure codes, through reabstraction of each sampled record by a professional coder and an obstetric nurse. These results are described in Section 2.

We received 1,614 of the 1,662 records that we requested from participating hospitals (97.1%). Three of these records did not actually represent deliveries, meaning that 99.8% of hospital-reported deliveries were confirmed as such. Reporting of cesarean versus vaginal delivery was also nearly perfect, with an unweighted sensitivity of 99.9% (e.g., one cesarean delivery was misreported as a vaginal delivery) and an unweighted positive predictive value (PPV) of 99.7% (e.g., two vaginal deliveries were misreported as cesarean deliveries). We conclude that the OSHPD's patient discharge data may be used to identify deliveries, and to distinguish cesarean from vaginal deliveries.

Question 2. What is the statewide reporting accuracy for risk factors in the risk adjustment models, using California patient discharge data?

Previous CHOP validation studies have demonstrated that the validity of risk factor data varies significantly, depending on the severity and importance of the risk factor. If certain risk factors are widely underreported or overreported by hospitals, then risk-adjustment models that include those factors may be biased. For example, if hospitals only report the most severe cases of a risk factor, then that risk factor may appear to have a much greater impact on patient outcomes than it actually does. To identify risk factors that may be coded too poorly to include in risk-adjustment models, professional coding experts completed a "blind" recoding of each discharge in the validation study. We compared their findings with the information that hospitals reported to the OSHPD, on a patient-by-patient basis. These results are described in Section 2.

Forceps and vacuum delivery were accurately reported, with sensitivities and predictive values exceeding 90%. Episiotomy was underreported (70% sensitivity), especially among women who experienced a third or fourth degree perineal laceration. All cesarean indications were reported with at least 60% sensitivity, except for uterine inertia, herpes, and long labor. Among comorbidities, sensitivity exceeded 60% for chorioamnionitis, diabetes, premature labor, preeclampsia, intrauterine death, and congenital abnormalities. Sensitivity was poor (<60%) for anemia, asthma, thyroid disorders, mental disorders, drug abuse, genitourinary infections, obesity, fibroids, excessive fetal growth, hypertension, premature rupture, polyhydramnios, and postdates. Based on these analyses of data from 1992 and 1993, we conclude that the OSHPD's patient discharge data may be used to ascertain the most important delivery-associated procedures and selected antepartum and intrapartum complications. However, certain risk factors cannot be validly ascertained, and should therefore be excluded from risk-adjustment models to reduce bias. Of course, we cannot exclude the possibility that the accuracy of reporting some of these risk factors may have improved statewide over the past decade.

Question 3. Is there a significant difference in the coding of important risk factors when comparing hospitals with significantly fewer readmissions than expected, significantly more readmissions than expected, and neither?

The ICD-9-CM coding of all key risk factors was examined by hospital outlier status, to determine whether variation in coding practices may explain why some hospitals appear to have better-than-expected outcomes and others appear to have worse-than-expected outcomes. In other words, hospitals that underreport risk factors may have relatively low expected readmission

rates, and thus relatively high risk-adjusted readmission rates, because their patients do not appear as sick as they actually are. To address this question, we compared the accuracy of risk factor reporting across hospitals stratified by whether they had fewer readmissions than expected, more readmissions than expected, or neither. These results are described in Section 2.

We found several statistically significant differences in reporting of risk factors between hospitals with more readmissions than expected and hospitals with fewer readmissions than expected, but no consistent pattern. Substituting recoded data for administrative data in our multivariate model to estimate readmission risk had minimal impact (e.g., hospitals that had 34% more readmissions than expected using administrative data still had 28% more readmissions than expected using recoded data, and hospitals that had 55% fewer readmissions than expected using administrative data still had 42% fewer readmissions than expected using recoded data). We conclude that differential underreporting of risk factors in the OSHPD's patient discharge data accounts for little of the observed variation in risk-adjusted readmission rates across hospitals.

Question 4. What is the statewide reporting accuracy for postpartum complications, using California patient discharge data?

Fortunately, maternal mortality is extremely rare in California. Therefore, consumers, purchasers, and providers of obstetric care are all interested in measures of potentially preventable postpartum complications. Several such measures have been endorsed by the Joint Commission on the Accreditation of Healthcare Organizations and the Agency for Healthcare Research and Quality. However, previous CHOP validation studies have shown that complications after back surgery are markedly underreported, and that variation in reporting practices accounts for a substantial proportion of the observed variation in complication rates between hospitals with fewer complications than expected and hospitals with more complications than expected. Do these findings apply to obstetric patients? To address this question, we again asked professional coding experts to complete a "blind" recoding of each discharge in the validation study, and we then compared their findings with the information that hospitals reported to the OSHPD. These results are described in Section 3.

Both third and fourth degree perineal lacerations were reported very accurately, with estimated sensitivities exceeding 90% and positive predictive values (PPVs) exceeding 65% (weighted) or 85% (unweighted). After in-depth review of discrepant cases, we estimate the actual PPV at over 90%. Most coding discrepancies for perineal laceration were between no injury and first degree, or between first and second degree. Most postpartum complications, including urinary tract and wound infections, endometritis, anesthesia complications, and postpartum hemorrhage were reported with less than 70% sensitivity, but at least 80% PPV. Composite measures from HealthGrades and Solucient, which include these complication codes, also suffer from high false negative rates. Based on these analyses of data from 1992 and 1993, we conclude that perineal lacerations are the only significant complication of inpatient obstetric care that can be validly ascertained using the OSHPD's patient discharge data. Of course, we cannot exclude the possibility that the accuracy of reporting other complications may have improved statewide over the past decade.

Question 5. Can vital statistics data (e.g., birth certificates) be used in place of, or in addition to, hospital discharge abstracts to ascertain maternal risk factors and adverse outcomes?

In the decade since the obstetric validation study was conceived, the OSHPD has begun routinely linking hospital discharge abstracts for new mothers and newborn babies with the corresponding birth certificate(s) in California's vital statistics registration database. Through this linkage, it may be possible to ascertain risk factors and potentially preventable complications more accurately than would be possible using hospital discharge abstracts alone. To use these linked data responsibly, one must understand when both data sets should be used to ascertain a risk factor of interest, and when one data set should be used in preference to the other. We found many previous studies evaluating the accuracy of data elements on birth certificates in the US, so we did not undertake any independent validation of birth certificate data. The published studies, and their implications for hospital quality measurement, are described in Section 4.

We found 484 unique citations and selected 100 papers for review, plus two that were identified through extended search methods. Studies of parental race/ethnicity and age (or date of birth) have reported excellent validity of birth certificate data. The accuracy of these data on parental education and occupation is lower, but at least 69%. Gravidity and parity, especially nulliparity, are accurately reported. Agreement between birth certificates and other data is moderate for timing of the last menstrual period and length of gestation. Almost all pregnancy-related conditions are poorly reported; prior cesarean is the only consistent exception. Birth certificates tend to exaggerate both the duration of prenatal care and the number of visits. Most complications of labor and delivery are reported with moderate sensitivity (20-75%). Method of delivery is accurately reported, but vaginal birth after cesarean (VBAC) is consistently underreported, with sensitivities of 39-70%. Reporting of forceps or vacuum usage has 60-83% sensitivity. Agreement between birth certificates and other data is high (≥93%) for birthweight, Apgar score, birth order, and infant gender. However, abnormal conditions of the newborn and congenital anomalies are consistently underreported, with sensitivities below 50% for most. We conclude that birth certificate data can be used to complement the OSHPD's patient discharge data by enhancing ascertainment of key ICD-9-CM coded risk factors and by identifying risk factors for which no ICD-9-CM codes exist (e.g., gravidity and parity). In general, birth certificates should not be relied upon as the sole source of data on antepartum and intrapartum risk factors.

Question 6. How does the risk-adjustment model change when additional clinical variables are used as risk factors?

Administrative datasets provide limited data, based on ICD-9-CM codes, to characterize patients' risk of readmission and other adverse outcomes. To assess how much risk-adjustment models using the OSHPD's data could be improved through supplementation with more detailed clinical data, our nurse abstractors collected key clinical data elements from all sampled medical records. These additional clinical variables were identified through a literature review and through discussions with a Clinical Advisory Panel. This information was then used, in part, to determine whether more complete clinical information would improve the validity of our risk-adjustment models for postpartum readmissions. These results are described in Section 5.

Adding several clinical data elements abstracted from medical records modestly improved the predictive performance of risk-adjustment models for postpartum readmission, at both the patient and hospital levels. These clinical data elements include the number of prior vaginal deliveries, body mass index (a measure of obesity), fetal macrosomia, antepartum fever, and thick meconium. A few other "clinical" risk factors were also independent predictors of postpartum readmission, but only because of underreporting in the OSHPD database. If these underreported risk factors, including prior cesarean, multiple gestation, insulin-requiring diabetes, psychiatric disorders, preterm gestation, preeclampsia or eclampsia, antepartum hemorrhage, and antepartum anemia, were reported with 100% sensitivity, then clinical abstraction would not have improved the risk-adjustment models. Overall, adding information from clinical abstraction of medical records improved the discrimination of predictive models for readmission from c=0.587 to c=0.619 after vaginal delivery, and from c=0.631 to c=0.651 after cesarean delivery. Hospitals that had 34% more readmissions than expected using administrative data still had 36% more readmissions than expected using additional clinical data, and hospitals that had 55% fewer readmissions than expected using administrative data still had 49% fewer readmissions than expected using additional clinical data. We were not able to evaluate whether using birth certificate diagnoses in conjunction with hospital discharge diagnoses would have improved ascertainment of these risk factors to the same extent as clinical abstraction. We conclude that analyses of risk-adjusted postpartum readmissions based on patient discharge data are not biased by unreported clinical risk factors that are documented in medical records.

Question 7. Do women who are delivered at hospitals with more readmissions than expected experience more postpartum complications than women who are delivered at hospitals with fewer readmissions than expected?

If risk-adjusted postpartum readmission rates are a valid quality indicator, then hospitals with high rates should also have high rates of in-hospital postpartum complications that are likely to worsen after discharge, thereby necessitating readmission. To address this question, our professional coders and nurse abstractors collected detailed information on postpartum complications by reviewing all sampled medical records. In the absence of such detailed review, hospitals with more readmissions than expected may appear to have excess complications, simply because of more conscientious reporting of adverse outcomes. These results are described in Section 5.

Most post-cesarean complications were more prevalent at high-readmission than at low-readmission hospitals (laceration, 9.4% versus 0.7%; wound infection, 5.9% versus 0.0%; endometritis, 13.3% versus 2.5%; estimated blood loss, 832 cc versus 648 cc; mean hematocrit drop 4.8% versus 3.9%). In addition, mean postpartum length-of-stay was shorter at low-readmission hospitals than elsewhere (vaginal, 28 versus 31-32 hours; cesarean, 70 versus 75-82 hours). We conclude that hospitals with more readmissions than expected actually did experience more adverse outcomes after delivery, relative to hospitals with fewer readmissions than expected.

Question 8. Are there meaningful differences in the process of care between hospitals with fewer readmissions than expected and hospitals with more readmissions than expected?

If risk-adjusted postpartum readmission rates are a valid quality indicator, then hospitals with high rates should manage labor, delivery, and the puerperium in ways that predispose to complications necessitating readmission. These management practices are also known as processes of care, because they describe the process by which nurses, physicians, and other health professionals provide care at the bedside. To address this question, we collected detailed information about medical and nursing management, although many important components of peripartum care could not be ascertained from medical records. For example, we hypothesized that several physician and nurse behaviors known to increase the risk of postpartum infection would be more prevalent at hospitals with high risk-adjusted readmission rates than at hospitals with low risk-adjusted readmission rates. These results are described in Section 5.

There were no significant differences in labor management, but women at high-readmission hospitals were less likely to have operative vaginal delivery (1.5% versus 16%) or scheduled repeat cesarean (17% versus 58%), and more likely to have vaginal birth after cesarean (60% versus 28%), than women at low-readmission hospitals. The latter difference was largely attributable to the fact that 16% of eligible patients at hospitals with fewer readmissions than expected were not counseled about a trial of labor, and 47% reportedly refused despite counseling. None of several physician and nurse behaviors that may increase the risk of postpartum infection was more prevalent at high-readmission hospitals than at low-readmission hospitals. We conclude that hospitals with fewer readmissions than expected generally practiced a more aggressive style of obstetric management than hospitals with more readmissions than expected, but we found no clear explanation for the observed difference in patient outcomes.

Question 9. Do hospitals with more readmissions than expected simply have a lower threshold for readmission than hospitals with fewer readmissions than expected?

If risk-adjusted postpartum readmission rates are a valid quality indicator, then patients who are readmitted after delivery at hospitals with high rates should be just as sick as patients who are readmitted after delivery elsewhere. Differences in severity of illness at readmission would suggest that hospitals with high readmission rates may be less selective in whom they readmit, or may see more vulnerable patients who would have difficulty managing their complications in the outpatient setting. To address this question, we collected detailed information from patients' histories, physical examinations, and laboratory studies about their severity of illness at the time of readmission. We then characterized each woman who was readmitted as meeting definite clinical criteria for readmission (i.e., readmission is the standard of care), any clinical criteria for readmission (i.e., readmission is acceptable but not essential), or no clinical criteria for readmission (i.e., readmission is probably unnecessary). These results are described in Section 5.

The percentage of women in the highest acuity level at readmission was similar at low (62%), intermediate (64%), and high-readmission (52%) hospitals. Only for postpartum endometritis was there evidence that women readmitted at high-readmission hospitals were less severely ill than women readmitted at low or intermediate-readmission hospitals (i.e., 42% met definite criteria for readmission, versus 56% and 62%, respectively). This difference disappeared when looser criteria for readmission were applied or when postpartum endometritis was aggregated with other complications that often necessitate readmission. We conclude that variation across

hospitals in the threshold for readmission may explain at least some of the difference in risk-adjusted readmission rates that is attributable to endometritis, but none of the difference that is attributable to postpartum hemorrhage, wound infection, or urinary tract infection.

Section 2: The Accuracy of Obstetric Diagnoses and Procedures Reported in California Patient Discharge Data

SUMMARY

Objective: To assess the validity of obstetric procedures and diagnoses in the California Patient Discharge Data Set.

Methods: We randomly sampled 1,611 deliveries from 52 of the 267 hospitals that performed more than 678 eligible deliveries in California in 1992-1993. We compared hospital-reported procedures and diagnoses against our recoding of the same records.

Results: Cesarean, forceps, and vacuum delivery were accurately reported, with sensitivities and predictive values exceeding 90%. Episiotomy was underreported (70% sensitivity). Cesarean indications were reported with at least 60% sensitivity, except for uterine inertia, herpes, and long labor. Among comorbidities, sensitivity exceeded 60% for chorioamnionitis, diabetes, premature labor, preeclampsia, intrauterine death, and congenital abnormalities. Sensitivity was poor (<60%) for anemia, asthma, thyroid disorders, mental disorders, drug abuse, genitourinary infections, obesity, fibroids, excessive fetal growth, hypertension, premature rupture, polyhydramnios, and postdates.

Conclusions: The validity of obstetric procedures and diagnoses on hospital discharge abstracts varies, with moderate-high accuracy for the most important codes.

The key results from this section have also been published in:

Yasmeen S, Romano PS, Schembri ME, Keyzer JM, Gilbert WM. Accuracy of obstetric diagnoses and procedures in hospital discharge data. *American Journal of Obstetrics and Gynecology* 2006;194(4):992-1001.

INTRODUCTION

Maternal hospital discharge abstracts have been used extensively to monitor trends in pregnancy-related treatments and outcomes such as cesarean births, [1] operative vaginal deliveries,[2] hospital length of stay and charges,[3] and perinatal morbidity and mortality.[4,5,6] The validity of these analyses depends upon the accuracy of the available data.

Multiple previous studies have evaluated the accuracy of diagnosis and procedure codes in Medicare claims data,[7,8] the Veterans Health Administration's Patient Treatment File,[9,10] the OSHPD Patient Discharge Data Set,[11,12,13] and similar data sets from individual hospitals[14] or other countries.[15] However, these studies have focused almost entirely on the general medical-surgical population.

Our objective was to validate coding of obstetric procedures and diagnoses in the nation's largest statewide discharge data program. We examined the sensitivity and positive predictive value of maternal hospital discharge abstracts, using the complete inpatient medical record as a gold standard. We hypothesized that clearly defined and clinically serious conditions such as placenta previa and malpresentation would be accurately coded, whereas more ambiguous conditions such as anemia and fetal distress would be poorly coded.

METHODS

We searched the MEDLINE database from 1985 through 2000 to identify clinical trials and case series reporting maternal outcomes of delivery. Additional papers were identified by a clinical advisory panel (which included four obstetricians and/or perinatologists, two family physicians, one obstetric nurse specialist, and one health information professional) and by reviewing reference lists in obstetrics texts and meta-analyses. We excluded papers without abstracts or in languages other than English, studies from developing countries, and studies limited to patients with unusual procedures or risk factors. We then reviewed abstracts to locate studies with at least 250 patients that identified risk factors for adverse postpartum outcomes. After discussing these findings with our advisory panel, we developed a comprehensive list of clinical risk factors, which we mapped to ICD-9-CM using appropriate references[16] with the assistance of two coding professionals. We also reviewed how the Joint Commission for the Accreditation of Healthcare Organizations (JCAHO) defines clinical risk factors for its Core Measures on Pregnancy and Related Complications.[17] Some risk factors were redefined or aggregated to capture differences in risk more precisely, based on the cesarean delivery rate and postpartum readmission rate associated with each ICD-9-CM diagnosis.

Inclusion and Exclusion Criteria

This retrospective cohort study was based on a stratified cluster sample of women between 10 and 55 years of age who were discharged from a nonfederal licensed acute care hospital in California, after giving birth, between January 1, 1992 and December 31, 1993. We defined delivery based on a pregnancy-related principal or secondary diagnosis of 640-676, with a fifth digit of 1 or 2, or 650 ("delivery in a completely normal case"). We excluded cases with a principal diagnosis of postpartum care (V24.x), hydatidiform mole (630), other abnormal product

of conception (631), or ectopic pregnancy (633.xx). We also excluded cases with a principal or secondary diagnosis of malignancy (141.x-172.x, 174.x-208.xx), missed abortion (632), or pregnancy with abortive outcome (634.xx-639.x). Finally, we excluded cases with a principal or secondary diagnosis of significant trauma (800.xx-839.xx, 850.xx-904.xx, 925.x-929.x, 940.x-958.x) or fetal death (656.4x, V27.1, V27.3-V27.4, V27.6-V27.7), if an external cause of injury code indicated a cause other than poisoning (E800-848, E880-899, E905-909, E916-926, E928, E950-958, E960-966, E968, E970-976, E980-988), with no suggestion of iatrogenic injury (E849.7, E870-876).

Cesarean delivery was defined to include all 74.xx codes except 74.3 (removal of extratubal ectopic pregnancy) and 74.91 (hysterotomy to terminate pregnancy, if not associated with a live birth [V27.0, V27.2-V27.3, V27.5-V27.6]). Vaginal deliveries were defined by default as all other deliveries, except that those with a principal or secondary diagnosis of "cesarean delivery without mention of indication" (669.7x) or a procedure code suggesting surgical delivery (74.3, 74.91) were excluded.

Linkage of Prior and Subsequent Hospitalizations

We linked delivery records with both postpartum (within 6 weeks after delivery) and antepartum hospitalizations (within 273 days before delivery), using the patient's SSN and date of birth. If two records had the same SSN but different dates of birth, then both were discarded to minimize the risk of false linkage. About 22% of vaginal and 15% of cesarean deliveries were excluded because of missing SSNs. To ensure a full period for ascertaining postpartum hospitalizations, we also excluded cases with delivery dates after November 19, 1993. The date of delivery was assigned as the date of the earliest delivery-associated procedure (72.xx, 73.5x-74.xx) or the date of admission (in the absence of any valid delivery-associated procedure dates). Our algorithm for defining readmissions is described more fully in Section 5 of this report.

Through this linkage process, we identified and reconciled cases that appeared to have two or more deliveries within 182 days; delivery records sharing the same admission date; records with admission dates 1-7 days apart with identical procedures and diagnoses; and records with overlapping admission and discharge dates. Most cases of the first anomaly were corrected by searching for specific coding errors (e.g., reporting a delivery diagnosis without a delivery procedure or outcome of delivery); uncorrectable cases were discarded. The same correction algorithm was applied to women with reported interpartum intervals of 182 to 223 days, except that uncorrectable cases were retained. Deliveries that occurred within six months after a molar (630-631), ectopic (633.x), or aborted (634.xx-637.xx, 639.x) pregnancy were excluded, unless retention of at least one viable fetus (651.3x-651.6x) was documented. The last three anomalies were resolved by manually selecting the more complete record of the same hospitalization, or by randomly selecting from among identical records; paired records that could not be "unduplicated" were discarded.

Sampling Hospitals and Cases

To ensure adequate representation, we identified and oversampled hospitals with fewer or more readmissions than expected. We did so by first excluding hospitals with no licensed perinatal

beds (15 in 1992), based on annual reports submitted to the OSHPD. We then used multivariate probit regression to estimate each patient's risk of experiencing a postpartum readmission. To develop these models, we split cases randomly, after stratifying by readmission status, into 60% estimation and 40% validation samples. We randomly generated ten 25% subsamples without replacement from the estimation sample. In each subsample, we used stepwise forward selection (p-to-enter <0.10) to identify independent predictors of readmission, after forcing in age, race, and a transformed instrumental variable representing the likelihood of cesarean delivery (derived from a separate model). We did not adjust for the actual method of delivery, because it may reflect quality of care (e.g., endogeneity bias).[18]

Candidate risk factors were defined as characteristics or conditions that probably existed at admission and that may have influenced patient outcomes. Hospitalization-related risk factors included source of payment, source of admission, and number of antepartum admissions (truncated at seven). Source of payment (uninsured, private, public) was used as a crude indicator of socioeconomic status, whereas source of admission (acute, skilled nursing, or intermediate care; all others) was a marker for high-risk patients who were referred to a regional center. Clinical risk factors were identified through the process described above, except that risk factors were dropped if they were extremely rare (<0.1%) or were not associated (p>0.10) with postpartum readmission.

Predictors that entered at least 5 of 10 subsample models were tested in the entire estimation sample. We used simple two-variable models to screen all two-way interactions involving the selected main effects, if present in at least 20 patients with readmissions. The final set of covariates was tested for robustness by comparing the parameter estimates from our estimation sample to the corresponding estimates derived by fitting the same model to our validation sample. After confirming robustness, we reestimated our model on the entire sample to generate more reliable parameter estimates. The resulting model had a c statistic of 0.630 and a barely significant Hosmer-Lemeshow statistic of χ^2=18.30 (p=0.019).[19]

In the first stage of sampling, we stratified the 267 eligible hospitals that performed more than 678 deliveries in the study period according to the number of postpartum readmissions: significantly (p<0.01) or marginally (0.01<p<0.10) more than expected, significantly or marginally fewer than expected, and none of the above. Within three of these five strata, we substratified northern California Kaiser hospitals to support a separate collaborative project. We randomly sampled 46 hospitals from the eight resulting strata (e.g., up to six per stratum). Five hospitals declined to participate and were replaced by ten randomly selected alternates. The resulting sample of 52 hospitals is representative of all acute care hospitals with active obstetric services in California: 3 city/county, 7 district, 10 Kaiser, 23 other non-profit, 1 private university, and 8 for-profit hospitals.

Next, we randomly sampled eligible patients within each sampled hospital. Records with one or more readmissions, and cesarean deliveries, were oversampled to boost the number of patients with adverse outcomes and thereby improve efficiency. Stratified random sampling increases the reliability of estimates for subsets of particular interest, but allows the researcher to generate unbiased population estimates using sampling weights, as described below. The statewide readmission rate during the study period was 1.0% for cesarean and 0.6% for vaginal deliveries;

the target readmission rate in our validation sample was 30%. Accordingly, we drew a sample of 1,662 deliveries, which were associated with 493 postpartum readmissions. The number of cases contributed by each hospital within a stratum was proportional to its total volume. The sample was designed to provide 80% power to detect a 20% absolute interstratum difference (e.g., 60% versus 80%) in sensitivity or positive predictive value (PPV) for a high-prevalence (e.g., 14%) condition, with a type I error rate of 5%.

Data Collection and Analysis

We asked each participating hospital to photocopy each sampled record, including associated prenatal records if available. Each record was reviewed by one of four experienced Accredited Record Technicians or Certified Coding Specialists, who recoded the ICD-9-CM diagnosis and procedure codes, as well as maternal demographic and prenatal data, blinded to the original discharge abstract. A regional coding authority tested these individuals before they were hired, trained and supervised them, and verified at least 10% of each abstractor's records to ensure at least 95% accuracy. Discrepancies were resolved through collective review of appropriate coding references.

Because of the complex sample structure, all analyses were weighted (unless otherwise noted). The weight was defined as the inverse of the sampling probability, which was calculated by multiplying the probability of sampling a specific hospital by the probability of sampling an individual within that hospital. These weights were adjusted to reflect both nonsubmitted records and records that were later classified as ineligible.

In this Section, we evaluate whether specific procedures and conditions can be accurately ascertained from the ICD-9-CM codes on hospital discharge abstracts. Accuracy was measured in terms of sensitivity and PPV, using our recoding of hospital charts as the gold standard. We defined sensitivity as the proportion of patients with a condition identified through recoding for whom the same condition was reported on the hospital's original discharge abstract. We defined PPV as the proportion of patients with a condition reported on the hospital's original discharge abstract for whom the same condition was independently found through recoding. We do not report specificity, or the percentage of patients without a complication (according to recoding) who were correctly reported as not having it, because this parameter was never below 95%, and nearly always exceeded 98%. Sensitivities were compared across hospital sampling strata using the svytab procedure in STATA Release 6, which provides robust variance estimates accounting for oversampling of cesarean deliveries and clustering of observations within hospitals.

The study protocol was approved by the appropriate committees at the University of California, Davis and the California Health and Human Services Agency.

RESULTS

We received 1,614 of the 1,662 records that we requested from participating hospitals (97.1%). Three of these records did not actually represent deliveries; 1,611 records were abstracted (30.3% primary cesarean, 18.9% repeat cesarean, 51.0% vaginal). This cohort had a weighted mean (SD) age of 28.0 (0.52) years, and a racial/ethnic composition similar to the target

population (55% white, 8% African American, 8% Asian, 29% Hispanic, 0.1% Native American, and 0.8% "other").

Tables 1 through 3 summarize the sensitivity and PPV of obstetric procedures and conditions that were documented in at least 12 women. Unweighted sensitivities and PPVs permit computation of the number of false positive and false negative cases. Weighted values permit extrapolation to the entire population of women who were delivered at acute care, non-federal hospitals with active obstetric services in California during the study period. However, weighted estimates are less stable because women who were readmitted had a higher sampling probability than women who were not. Therefore, the weighted estimate should be interpreted cautiously when it differs greatly from the corresponding unweighted estimate. Unless otherwise stated, weighted estimates are cited below.

Table 1 shows that cesarean, forceps, and vacuum deliveries were reported very accurately, with sensitivities exceeding 89% and PPVs exceeding 96%. When 19 of the 20 discrepant cases were reviewed by the lead nurse abstractor and the principal investigator, 6 were resolved in favor of the original discharge abstract and 13 were resolved in favor of our recoding. Episiotomy was underreported, with a sensitivity of 70%, but rarely overreported, with a PPV of 95%. A thorough review of discrepant cases confirmed our recoding in all but one case of suspected underreporting, but 6 of the 14 cases of suspected overreporting were found to be true positives. Induction of labor was highly underreported, with sensitivities of 32-45%, and modestly overreported, with PPVs of 76-88%.

Table 2 shows moderately high concordance between hospital-reported and recoded data for most potential indications for cesarean delivery. Both sensitivity and PPV (either weighted or unweighted) exceeded 85% for malpresentation, previous cesarean, shoulder dystocia, multiple gestation, occiput posterior, placenta previa, and abruption. Uterine inertia, obstructed labor, genital herpes, long labor, antepartum hemorrhage, and failed induction were substantially underreported, with sensitivities of 60% or less. Long labor was substantially overreported, with PPV below 60%.

Table 3 shows variable concordance between hospital-reported and recoded data for other pregnancy-related conditions. Both sensitivity and PPV (either weighted or unweighted) exceeded 80% only for preeclampsia, intrauterine death, and congenital uterine abnormalities. Among the other conditions evaluated, only chorioamnionitis (79% sensitivity, 87% PPV), diabetes (64% sensitivity, 96% PPV), premature labor (77% sensitivity, 96% PPV), and prolonged rupture (65% sensitivity, 66% PPV) approached these thresholds. Most of the other pregnancy-related conditions shown in Table 3 were markedly underreported.

In secondary analyses (not shown), we evaluated the impact of defining conditions based on both delivery records and linked antepartum records. In no case did the sensitivity of reporting increase by more than 5%. To determine whether hospitals with high risk-adjusted readmission rates actually had sicker patients than their reported data suggested, we compared the sensitivity of reporting across sampling strata. We found several statistically significant differences, as shown in Table 4, but no consistent pattern. Substituting recoded data for administrative data in our multivariate model to estimate readmission risk had minimal impact (e.g., hospitals that had

34% more readmissions than expected using administrative data still had 28% more readmissions than expected using recoded data, and hospitals that had 55% fewer readmissions than expected using administrative data still had 42% fewer readmissions than expected using recoded data). Finally, we explored whether the sensitivity of reporting for episiotomy was related to the occurrence of perineal injury. About 62% of episiotomies were reported among women with third or fourth degree lacerations, compared with 71% among women without such injuries.

Finally, Table 5 provides a comprehensive summary of the sensitivity and PPV of every risk factor that JCAHO uses in risk-adjustment of its Core Measures on Pregnancy and Related Complications. In some cases, the JCAHO definition of a risk factor, shown in this table, differs from our preferred definition of the same risk factor, shown in Tables 2 or 3.

DISCUSSION

The current study represents a comprehensive analysis of the validity of obstetric diagnosis and procedure codes on 1,611 discharge abstracts from 52 California hospitals. Validity varied widely across procedures and conditions. Surgical procedures such as cesarean and instrumented vaginal delivery were accurately reported, with both sensitivities and PPVs exceeding 90%. Episiotomy was an exception, with 30% underreporting in the hospital database. Underreporting was particularly frequent among women with high-grade lacerations, because official coding policy states that when "an episiotomy extends spontaneously to become a perineal laceration… the laceration is coded as a diagnosis and no code is assigned for the episiotomy."[16] Induction of labor was markedly underreported, consistent with the Uniform Hospital Discharge Data Set's definition of a "significant" procedure as one that is "surgical in nature, or carries a procedural risk, or carries an anesthetic risk, or requires specialized training." False positive errors were generally attributable to confusion between inducing and augmenting labor.

Most potential indications for cesarean delivery were reported with moderate or high accuracy, with the notable exceptions of uterine inertia, obstructed labor, genital herpes, long labor, and antepartum hemorrhage. These conditions may be documented more clearly in the medical record, and hence coded more often, when they represent the actual indication for cesarean delivery. Hospitals are only required to report "conditions that affect patient care in terms of requiring: clinical evaluation; or therapeutic treatment; or diagnostic procedures; or extended length of hospital stay; or increased nursing care and/or monitoring."[20] According to ICD-9-CM Official Guidelines, "diagnoses that relate to an earlier episode which have no bearing on the current hospital stay are to be excluded." Hence, genital herpes must be reported only if it is active at the time of delivery and it affects obstetric management.

Among various pregnancy-related comorbidities, the sensitivity of reporting was moderate (60-80%) or high (over 80%) for chorioamnionitis, diabetes, preeclampsia, premature labor, intrauterine death, congenital uterine abnormalities, and prolonged rupture. The sensitivity of reporting was poor (40-60%) or very poor (<40%) for all other comorbidities. False positive errors were generally less frequent than false negative errors, and some may represent diagnoses missed on recoding. Most false positives for anemia, genitourinary infections, and obesity appear attributable to hospitals' improperly coding from laboratory or physical findings rather than from physician notes. *Coding Clinic for ICD-9-CM* has consistently advised coders not to

"reach into the medical record to code other conditions for the sake of coding...(if in doubt) contact the physician."[21] Therefore, a positive urine culture or a low hematocrit is not codable unless the treating physician documents a codable diagnosis.

Our findings suggest that computerized hospital discharge abstracts are a valid source of data on delivery type, multiple births, malpresentation, and selected pregnancy-related conditions such as chorioamnionitis and preeclampsia. The validity of data on many other maternal risk factors may be adequate for some applications, but not for others. For example, estimates of the association between episiotomy and perineal laceration based on hospital discharge data are probably biased toward the null,[22] because an episiotomy that extends into a 3rd or 4th degree laceration need not be reported. Conversely, estimates of the association between ill-defined (e.g., fetal distress) or mutable (e.g., herpes) risk factors and cesarean delivery are probably biased away from the null, because these risk factors must be reported only if they affect obstetric management. Although reporting practices vary across hospitals, we saw no evidence of systematic bias in estimating risk-adjusted readmission rates from discharge data.

Our results are generally consistent with the few prior studies on the validity of obstetric procedures and diagnoses in hospital administrative datasets. Using linked discharge abstracts and birth certificates as a limited gold standard, Keeler and colleagues reported sensitivities of 99% for cesarean delivery, 89% for prior cesarean, and 91% for breech presentation in 1989-1990 Washington discharge data.[23] Parrish and colleagues independently reabstracted 7,536 of these records and reported sensitivities of 98% for cesarean delivery, 78% for forceps or vacuum delivery, 56% for episiotomy, 95% for prior cesarean in the setting of a repeat cesarean, 68% for prior cesarean with vaginal delivery, 87% for fetal distress, and 81% for disproportion.[24] In a smaller study from one center,[25] Korst and colleagues reported sensitivities of 81% for malpresentation, 20% for "antepartum bleeding or placental conditions" (including previa, abruption, and other/unspecified antepartum hemorrhage), 32% for herpes, 100% for multiple gestation, 50% for excessive fetal growth, 44% for "soft tissue conditions" (including congenital uterine abnormalities, fibroids, cervical incompetence, and other pelvic floor problems), 100% for severe preeclampsia or eclampsia, 70% for other preeclampsia or hypertension, and 84% for early onset of delivery. PPVs exceeded 80% for all conditions except excessive fetal growth (35%) and "soft tissue conditions" (64%).

The most important limitation of this research is that our recoded data do not provide an ideal gold standard. However, we selected coders with experience in obstetrics, trained them thoroughly using specific written guidelines, monitored them carefully, and gave them unlimited time to abstract each record. When cases with discrepant procedural data were thoroughly reviewed by the lead nurse abstractor and the principal investigator, nearly every false negative was confirmed but some false positives were found to represent recoding errors. Second, because this study was designed to validate a published report on risk-adjusted outcomes, our medical records came from 1992 and 1993. ICD-9-CM coding of obstetric diagnoses may have improved in the past decade. The mean number of diagnoses reported on all California hospital discharge abstracts increased from 4.50 in 1997 to 5.46 in 2003,[26] although comparable statistics are not available for obstetric abstracts. Third, linked databases that include both maternal discharge abstracts and newborn birth certificates may allow researchers to overcome underreporting in hospital data. For example, using the linked database instead of the OSHPD

hospital discharge database to ascertain perinatal risk factors in California in 2000-2002, estimated prevalence increased by 7% for abruption, 8% for diabetes, 22% for premature labor, 3% for hypertension, 26-29% for preeclampsia, 32% for breech, 34% for long labor, 43% for induction of labor, 4% for chorioamnionitis, 16% for anemia, 40% for placenta previa, and 11% for cardiac disease. Finally, our exclusion of very low-volume hospitals from the sampling frame may limit the generalizability of our findings. Fortunately, these hospitals accounted for fewer than 10% of deliveries performed in California hospitals during the study period.

Despite these limitations, our findings have important implications for epidemiologic and health policy research. Hospital discharge data are very useful for studying delivery-related procedures, most indications for cesarean delivery, and a few other perinatal risk factors. However, they seem less useful for identifying chronic maternal conditions (e.g., asthma, thyroid disorders, alcohol or drug abuse, obesity), ill-defined conditions of uncertain significance (e.g., uterine inertia, fetal distress, long labor), and antepartum conditions that may resolve before delivery (e.g., anemia, genitourinary infections). Researchers should avoid relying on hospital discharge data alone to ascertain these diagnoses. Health information professionals should direct more attention to coding obstetric diagnoses, which rarely affect the Medicare population and have therefore been excluded from previous national studies of coding and DRG accuracy. Given that obstetric diagnoses usually do not affect hospital or physician reimbursement, obstetric records may not be abstracted as carefully as medical-surgical records. For the same reason, physicians providing obstetric care may not face as much local scrutiny of their documentation practices. Obstetric diagnoses are important for quality monitoring and for clinical and epidemiologic research, even if there are no national audits and no financial incentives to code them accurately. Finally, several ICD-9-CM obstetric codes would benefit from clearer definitions that are more consistent with the terminology currently used in the obstetric community.

REFERENCES

1. Gregory KD, Korst KM, Gornbein JA, Platt LD. Using administrative data to identify indications for elective primary cesarean delivery. *Health Serv Res* 2002;37:1387-1401.
2. Demissie K, Rhoads GG, Smulian JC, Balasubramanian BA, Gandhi K, Joseph KS, et al. Operative vaginal delivery and neonatal and infant adverse outcomes: population based retrospective analysis. *BMJ* 2004;329:1-6.
3. Garcia FAR, Miller HB, Huggins GR, Gordon TA. Effect of academic affiliation and obstetric volume on clinical outcome and cost of childbirth. *Obstet Gynecol* 2001;97:567-576.
4. Coronado GD, Marshall LM, Schwartz SM. Complications in pregnancy, labor, and delivery with uterine leiomyomas: A population-based study. *Obstet Gynecol* 2000;95:764-769.
5. Lydon-Rochelle M, Holt VL, Easterling TR, Martin DP. First-birth cesarean and placental abruption or previa at second birth. *Obstet Gynecol* 2001;97:765-769.
6. Gilbert WM, Nesbitt TS, Danielsen B. Childbearing beyond age 40: pregnancy outcome in 24,032 cases. *Obstet Gynecol* 1999;93:9-14.
7. Fisher ES, Whaley FS, Krushat WM, Malenka DJ, Fleming C, Baron JA, Hsia DC. The accuracy of Medicare's hospital claims data: Progress has been made, but problems remain. *Am J Public Health* 1992;82:243-248.
8. Hartz AJ, Kuhn EM. Comparing hospitals that peform coronary artery bypass surgery: The effect of outcome measures and data sources. *Am J Public Health* 1994;84:1609-1614.
9. Geraci JM, Ashton CM, Kuykendall DH, Johnson ML, Wu L. International classification of Diseases, 9th Revision, Clinical Modifications Codes in discharge abstracts are poor measures of complication occurrence in medical inpatients. *Med Care* 1997;35:589-602.
10. Best WR, Khuri SF, Phelan M, Hur K, Henderson WG, Demakis JG, et al. Identifying patient preoperative risk factors and postoperative adverse events in administrative databases: Results from the Department of Veterans Affairs National Surgical Quality Improvement Program. *J Am Coll Surg* 2002;194:257–266.
11. Romano PS, Mark DH. Bias in the coding of hospital discharge data and its implications for quality assessment. *Med Care* 1994;32:81-90.
12. Green J, Wintfeld N. How accurate are hospital discharge data for evaluating effectiveness of care? *Med Care* 1993;31(8):719-31.
13. Romano PS, Schembri ME, Rainwater JA. Can administrative data be used to ascertain clinically significant postoperative complications? *Am J Med Qual* 2002;17:145-154.
14. Jollis JG, Ancukiewicz M, DeLong ER, et al. Discordance of databases designed for claims payment versus clinical information systems. *Ann Intern Med* 1993;119:844–850.
15. Quan H, Parsons GA, Ghali WA. Validity of information on comorbidity derived from ICD-9-CCM administrative data. *Med Care* 2002; 40:675-685.
16. *American Hospital Association's ICD-9-CM Coding Handbook, 2004 Revised Edition*. American Hospital Association: Chicago, IL; 2004.
17. http://www.jcaho.org/pms/core+measures/appendixb.pdf , accessed March 18 2005.
18. Lydon-Rochelle M, Holt VL, Martin DP, et al. Association between method of delivery and maternal rehospitalization. *JAMA* 2000;283:2411-2416.
19. Romano PS, Luft HS, Remy LL, et al. *Report of the California Hospital Outcomes Project: Maternal Outcomes Following Delivery. Risk-adjustment Methodology and Preliminary Findings*. Sacramento, CA: Office of Statewide Health Planning and Development; 1996.
20. *Coding Clinic for ICD-9-CM* 1990 (2nd Quarter):24.

21. *Coding Clinic for ICD-9-CM* 1984 (May-June):12.
22. Handa VL, Danielsen BH, Gilbert WM. Obstetric anal sphincter lacerations. *Obstet Gynecol* 2001; 98(2):225-30.
23. Keeler EB, Park RE, Bell RM, Gifford DS, Keesey J. Adjusting cesarean delivery rates for case mix. *Health Serv Res* 1997;32:511-528.
24. Parrish KM, Holt VL, Connell FA, Williams B, LoGerfo JP. Variations in the accuracy of obstetric procedures and diagnoses on birth records in Washington State, 1989. *Am J Epidemiol* 1993; 138:119-27.
25. Korst LM, Gregory KD, Gornbein JA. Elective primary caesarean delivery: accuracy of administrative data. *Pediatr Perinat Epidemiol* 2004;18:112-119.
26. http://www.oshpd.ca.gov/HQAD/PatientLevel/ICD9Codes.htm, accessed December 6 2004.

Table 1. Validity of ICD-9-CM delivery-related procedures on California hospital discharge abstracts

Procedure	ICD-9-CM codes	No. (%) women with condition	Sensitivity (%)		Positive Predictive Value (%)	
			Unweighted	Weighted	Unweighted	Weighted
Cesarean delivery	74.0-74.2, 74.4, 74.99	789 (21.5%)	100%	100%	100%	100%
Induction of labor (any)	73.01, 73.1, 73.4, 659.0x-659.1x	244 (14.7%)	49%	45%	72%	88%
Medical induction (vaginal deliveries only)	73.4	110 (10.2%)	46%	42%	74%	84%
Surgical induction (vaginal deliveries only)	73.01, 73.1	32 (3.6%)	31%	32%	63%	76%
Forceps/Vacuum (vaginal deliveries only)	72.0-72.4, 72.51, 72.53, 72.6 (forceps)	38 (4.5%)	87%	89%	92%	99%
	72.7x (vacuum)	100 (10.1%)	90%	94%	98%	96%
	72.0-72.4, 72.51, 72.53, 72.6-72.9 (either)	131 (14.5%)	93%	95%	98%	98%
Episiotomy (vaginal deliveries only)	72.1, 72.21, 72.31, 72.71, 73.6	410 (53.9%)	72%	70%	95%	95%
Repair of obstetric laceration of rectum and sphincter ani (vaginal deliveries only)	75.62	54 (4.8%)	83%	51%	85%	41%

Table 2. Validity of ICD-9-CM conditions representing potential indications for cesarean delivery on California hospital discharge abstracts

Condition	ICD-9-CM codes	No. (%) women with condition	Sensitivity (%)		Positive Predictive Value (%)	
			Unweighted	Weighted	Unweighted	Weighted
Uterine inertia	661.0x- 661.2x	256 (7.2%)	59%	56%	82%	84%
Malpresentation, specified type	652.2x-652.4x, 652.6x, 669.6x	124 (3.9%)	90%	88%	85%	67%
Malpresentation, any	652.2x-652.4x, 652.6x-652.9x, 669.6x	192 (7.6%)	85%	90%	93%	97%
Cesarean, previous	654.2x	366 (14.8%)	87%	74%	98%	91%
Disproportion*	653.xx	205 (6.6%)	76%	64%	91%	92%
Obstructed labor (caused by maternal problem)	660.1x, 660.2x	183 (6.2%)	50%	40%	85%	80%
Shoulder dystocia*	660.4x	16 (1.6%)	81%	99%	93%	98%
Fetal distress*	656.3x	230 (9.7%)	66%	68%	80%	69%
Genital herpes	054.10-054.12, 054.19, 054.79-054.9	38 (2.3%)	29%	9%	73%	69%
Long labor*	662.0x-662.2x	24 (0.4%)	46%	36%	52%	25%
Multiple gestation	651.xx, 652.6x, 660.5x, 662.3x, V27.2-V27.7	44 (2.3%)	95%	92%	100%	100%
Occiput posterior	660.3x	48 (3.3%)	56%	60%	60%	81%
Placenta previa	641.0x, 641.1x	24 (0.8%)	79%	88%	100%	100%
Premature separation of placenta (abruptio)*	641.2x	27 (0.6%)	89%	63%	89%	82%
Antepartum hemorrhage*	641.1x, 641.3x, 641.8x, 641.9x	24 (0.7%)	54%	46%	72%	70%
Failed induction	659.0x, 659.1x	26 (1.6%)	42%	60%	28%	68%

* Used for risk-adjustment in the Joint Commission for the Accreditation of Healthcare Organizations' (JCAHO) Pregnancy and Related Complications Core Measures. The sensitivity and PPV of the JCAHO-defined risk factor are identical to the values shown.

Table 3. Validity of ICD-9-CM pregnancy-related comorbid conditions on California hospital discharge abstracts

Condition	ICD-9-CM codes	No. (%) women with condition	Sensitivity (%)		Positive Predictive Value (%)	
			Unweighted	Weighted	Unweighted	Weighted
Infection of amniotic cavity(chorioamnionitis)	658.4x, 659.3x	75 (2.3%)	75%	79%	86%	87%
Anemia (antepartum)	280.x-284.x, 285.0, 285.8[§], 648.21, 648.23	89 (4.4%)	30%	12%	28%	14%
Asthma	493.xx	22 (2.3%)	45%	42%	77%	91%
Diabetes mellitus	250.xx, 357.2, 362.0, 648.0x	14 (0.2%)	71%	75%	50%	23%
Diabetes mellitus or abnormal glucose tolerance (gestational)[*]	250.xx, 357.2, 362.0, 648.0x, 648.8x, 790.2	112 (6.0%)	78%	64%	94%	96%
Thyroid disorders	240.x-246.x, 648.1x	24 (2.8%)	33%	10%	89%	100%
Alcohol abuse and mental disorders	290.xx-291.xx, 294.xx-303.xx, 305.0x, 306.xx-319, 357.5, 425.5, 535.3x, 571.0-571.3, 648.41, 648.43, 980.0, 980.9, V11.3	176 (7.4%)	13%	15%	92%	97%
Drug abuse	292.0, 292.82, 304.xx, 305.2x-305.9x, 648.31, 648.33	51 (2.8%)	41%	38%	91%	98%
Tobacco use disorder[*]	305.1	149 (5.3%)	3%	15%	71%	96%
Genitourinary infections of pregnancy	646.61, 646.63	30 (1.2%)	40%	39%	43%	45%
Obesity	278.0[§]	65 (1.4%)	12%	11%	62%	49%
Excessive weight gain	646.1x	43 (1.8%)	5%	3%	29%	19%
Obesity or excessive weight gain[†]	278.0[§], 646.1x	105 (3.2%)	14%	13%	83%	71%
Preeclampsia, severe	642.5x, 642.7x	31 (0.8%)	84%	76%	84%	94%
Preeclampsia, any[†]	642.3x, 642.4x, 642.5x, 642.7x	71 (2.0%)	80%	88%	77%	91%

Early onset of delivery (premature labor)	644.2x	165 (8.0%)	75%	77%	95%	96%
Precipitate labor*	661.3x	39 (5.9%)	38%	31%	44%	45%
Intrauterine death	656.4x, V27.1, V27.3-V27.4, V27.6-V27.7	14 (0.9%)	86%	74%	100%	100%
Congenital uterine abnormality	654.0x, 752.2, 752.3	13 (1.2%)	54%	91%	100%	100%
Uterine fibroids	218.x, 654.1x	65 (1.5%)	40%	37%	100%	100%
Fetal growth, excessive (large-for-dates)*	656.6x	73 (3.0%)	45%	41%	72%	88%
Fetal growth, poor (small-for-dates)	656.5x	28 (1.6%)	68%	27%	95%	93%
Hypertension	401.xx-405.xx, 429.3, 642.0x-642.2x, 642.7x, 642.9x	80 (2.5%)	58%	58%	74%	86%
Premature rupture of membranes	658.1x	85 (3.1%)	54%	45%	64%	57%
Prolonged rupture (delayed delivery)	658.2x, 658.3x	46 (3.2%)	46%	65%	47%	66%
Oligohydramnios*	658.0x	53 (1.9%)	62%	55%	89%	96%
Polyhydramnios*	657.0x	14 (1.3%)	50%	14%	100%	100%
Postdates pregnancy	645.0x§	88 (4.2%)	48%	50%	74%	71%
Cardiovascular disease	393-398.xx, 648.5x, 648.61, 648.63, 745.xx-747.xx	37 (1.8%)	27%	12%	83%	99%
Elderly gravida	659.5x, 659.6x	104 (9.2%)	15%	3%	55%	32%

* Used for risk-adjustment in the Joint Commission for the Accreditation of Healthcare Organizations' (JCAHO) Pregnancy and Related Complications Core Measures. The sensitivity and PPV of the JCAHO-defined risk factor are identical to the values shown.
† An almost identical risk factor is used for risk-adjustment in the JCAHO Pregnancy and Related Complications Core Measure Set. The sensitivity and positive predictive value of the JCAHO-defined risk factor are within 1% of the values shown.
§ This ICD-9-CM code was correct at the time of delivery, but has been subsequently revised.

Table 4. Variability in the sensitivity of reporting potential indications for cesarean delivery and pregnancy-related comorbid conditions on California hospital discharge abstracts, across hospitals with more readmissions than expected, fewer readmissions than expected, and neither (only pairwise comparisons that differ at p<0.05 are shown; shaded cells indicate hospital stratum/strata with highest sensitivity)

Condition	**No. women with condition**	**Weighted Sensitivity (%)**			
		Overall (N=1,611)	**Hospitals with fewer (p<0.10) readmissions (N=726)**	**Other Hospitals (N=300)**	**Hospitals with more (p<0.10) readmissions (N=585)**
Malpresentation, any	192	90%	88%	96%[1]	73%[1]
Multiple gestation	44	92%	100%[3]	100%[1]	41%[3,1]
Occiput posterior	48	60%	62%[4]	58%	95%[4]
Placenta previa	24	88%	70%	100%[4]	71%[4]
Premature separation of placenta (abruption)	27	63%	88%[4]	0%[4,5]	39%[5]
Infection of amniotic cavity (chorioamnionitis)	75	79%	52%[2]	99%[2]	74%
Diabetes mellitus	14	75%	66%	47%[3]	100%[3]
Thyroid disorders	24	10%	4%[4]	0%[5]	75%[4,5]
Genitourinary infections of pregnancy	30	39%	21%[4]	99%[4,5]	33%[5]
Early onset of delivery (premature labor)	165	77%	75%	91%[6]	58%[6]
Intrauterine death	14	74%	100%[4]	0%[4,5]	68%[5]
Congenital uterine abnormality	13	91%	67%[5,3]	100%[3,5]	3%[5]
Uterine fibroids	65	37%	57%[4]	2%[4,1]	54%[1]
Fetal growth, poor (small-for-dates)	28	27%	73%[3]	13%[3]	41%
Premature rupture of membranes	85	45%	71%[1]	28%[1]	44%
Oligohydramnios	53	55%	74%[3]	49%	46%[3]
Polyhydramnios	14	14%	100%[4,5]	0%[4]	9%[5]
Cardiovascular disease	37	12%	42%[1]	0%[1]	12%
Elderly gravida	104	3%	17%[1]	0%[1,3]	6%[3]

[1] The two percentages in the same row that are marked with this superscript number are significantly different by pairwise comparison, $p<0.05$.
[2] The two percentages in the same row that are marked with this superscript number are significantly different by pairwise comparison, $p<0.005$.
[3] The two percentages in the same row that are marked with this superscript number are significantly different by pairwise comparison, $p<0.05$.
[4] The two percentages in the same row that are marked with this superscript number are significantly different by pairwise comparison, $p<0.0005$.
[5] Any pair of percentages in the same row that are marked with this superscript number are significantly different by pairwise comparison, $p<0.0005$.
[6] The two percentages in the same row that are marked with this superscript number are significantly different by pairwise comparison, $p<0.001$.

Table 5. Validity of ICD-9-CM conditions used by the Joint Commission on Accreditation of Healthcare Organizations in risk-adjustment for the Pregnancy and Related Complications Core Measure Set

Condition	ICD-9-CM codes	Women with condition		Sensitivity (%)		Positive Predictive Value (%)	
		No.	%	Unweighted	Weighted	Unweighted	Weighted
Diabetes in Pregnancy	648.01, 648.81	112	6.0%	78%	64%	95%	96%
Smoker	305.1	149	5.3%	3%	15%	71%	96%
Obesity	278.0, 278.01, 646.11	103	3.0%	15%	14%	83%	71%
Fetal Distress	656.31	230	9.7%	66%	68%	80%	69%
Abnormal Presentation (PR1)	652.01, 652.11, 652.21, 652.31, 652.41, 652.51, 652.61, 652.71, 652.81	234	8.8%	76%	82%	87%	90%
Multiple Gestations (PR1)	651.01, 651.11, 651.21, 651.31, 651.41, 651.51, 651.61, 651.81, 651.91, 659.41,V27.2, V27.3, V27.4, V27.5, V27.6, V27.7, V27.9	51	2.8%	82%	77%	93%	95%
Maternal Infections	646.51, 646.61, 647.81, 647.91, 658.41,659.31	113	3.7%	62%	64%	71%	75%
Cord Prolapse	663.01	9	0.1%	67%	59%	75%	91%
Obstructed Labor	660.01, 660.11, 660.21, 660.31, 660.41, 660.51, 660.61, 660.71, 660.81, 660.91	320	13.6%	53%	54%	73%	80%
Other Hypertensions in Pregnancy	642.01, 642.11, 642.21, 642.31, 642.91	71	2.1%	51%	60%	67%	84%
Polyhydraminios	657.01	14	1.3%	50%	14%	100%	100%
Abnormality of Organs and Soft Tissues of the Pelvis	654.01, 654.11, 654.31, 654.41, 654.51, 654.61, 654.71, 654.81, 654.91	177	5.4%	38%	42%	94%	95%
Large Fetus (Oversize)	656.61	73	3.0%	45%	41%	72%	88%
Eclampsia, Pre-eclampsia	642.41, 642.51, 642.61, 642.71	71	2.0%	79%	88%	76%	90%
Fetal-Maternal Hemorrhage	656.01	0	0.0%	0%	0%	0%	0%
Isoimmunization	656.11, 656.21	3	0.2%	67%	93%	4%	4%
Placenta Previa w/o Hemorrhage	641.01	4	0.2%	50%	84%	50%	53%

Abruptio Placenta	641.21	27	0.6%	89%	63%	89%	82%
Antepartum Hemorrhage	641.11, 641.31, 641.81, 641.91	24	0.7%	54%	46%	72%	70%
Disproportion	653.01, 653.11, 653.21, 653.31, 653.41, 653.51, 653.61, 653.71, 653.81, 653.91	205	6.6%	76%	64%	91%	92%
Failure to Progress	659.01, 659.11, 661.01, 661.41, 661.11, 661.21, 662.01, 662.11, 662.21	292	8.5%	65%	64%	83%	87%
Oligohydramnios	658.01	53	1.9%	62%	55%	89%	96%
Prolonged Labor	662.01, 662.11, 662.21	24	0.4%	46%	36%	52%	25%
Induction	73.1, 73.4	207	11.7%	42%	36%	65%	82%
Abnormal Presentation (PR3)	652.01, 652.11, 652.21, 652.31, 652.41, 652.51, 652.61, 652.71, 652.81	234	8.8%	76%	82%	87%	90%
Multiple Gestations (PR3)	651.01, 651.11, 651.21, 651.31, 651.41, 651.51, 651.61, 651.81, 651.91, 659.41, V27.2, V27.3, V27.4, V27.5, V27.6, V27.7, V27.9	51	2.8%	82%	77%	93%	95%
Disproportion	653.01, 653.11, 653.21, 653.31, 653.41, 653.51, 653.61, 653.71, 653.81, 653.91	205	6.6%	76%	64%	91%	92%
Large Fetus (Oversize) (PR3)	656.61	73	3.0%	45%	41%	72%	88%
Precipitate Labor	661.31	39	5.9%	38%	31%	44%	45%
Episiotomy	73.6	315	33.8%	64%	63%	93%	93%
Operative Vaginal Delivery not including breech, vacuum or Forceps	72.4, 72.51, 72.52, 72.53, 72.54, 72.6, 72.8, 72.9	9	2.3%	67%	87%	67%	97%
Breech Delivery	72.5	5	1.4%	60%	81%	75%	96%
Shoulder Dystocia	660.41	16	1.6%	81%	99%	93%	98%
Vacuum Extraction	72.7, 72.71, 72.79	109	8.0%	83%	94%	95%	91%
Forceps Delivery	72.0, 72.1, 72.2, 72.3, 72.21, 72.29,72.31, 72.39	38	3.5%	87%	89%	89%	94%

Section 3: Coding of Perineal Lacerations and Other Complications of Obstetric Care in California Patient Discharge Data

SUMMARY

Objective: To assess the validity of obstetric complications, including the Joint Commission on Accreditation of Healthcare Organizations (JCAHO) core measure on perineal lacerations, in the California Patient Discharge Data Set.

Methods: We randomly sampled 1,611 deliveries from 52 of the 267 hospitals that performed more than 678 eligible deliveries in California in 1992-1993. We compared hospital-reported complications against our recoding of the same records.

Results: Third and fourth degree perineal lacerations were reported accurately, with estimated sensitivities exceeding 90% and positive predictive values (PPV) exceeding 65% (weighted) or 85% (unweighted). Based on in-depth review of discrepant cases, we estimate the actual PPV at over 90%. Most coding discrepancies were between no injury and first degree, or between first and second degree. Most postpartum complications, including urinary tract and wound infections, endometritis, anesthesia complications, and postpartum hemorrhage were reported with less than 70% sensitivity, but at least 80% PPV. Composite measures from HealthGrades and Solucient, which include these complication codes, also suffer from high false negative rates.

Conclusions: Third and fourth degree perineal lacerations are accurately reported on hospital discharge abstracts, confirming the validity of related quality indicators sponsored by the Agency for Healthcare Research and Quality and JCAHO. Administrative data seem less useful for monitoring other in-hospital postpartum complications.

The key results from this section have also been published in:

Romano PS, Yasmeen S, Schembri ME, Keyzer JM, Gilbert WM. Coding of perineal lacerations and other complications of obstetric care in hospital discharge data. *Obstetrics and Gynecology* 2005;106(4):717-725.

INTRODUCTION

The growing interest in health care quality has stimulated efforts to measure hospital outcomes using secondary data sources such as birth certificates, hospital discharge abstracts, and insurance claims. For example, the Joint Commission on Accreditation of Health Care Organizations (JCAHO) has identified three Core Measures of pregnancy-related hospital care: vaginal birth after cesarean, inpatient neonatal mortality, and third or fourth degree laceration.[1] All of these measures are designed for use with hospital administrative data. More recently, the Agency for Healthcare Research and Quality (AHRQ), through its Quality Indicators project,[2] has promoted the use of ICD-9-CM coded data to ascertain obstetric trauma.[3,4] Laceration rates have also been used to monitor and improve quality at the local level.[5] HealthGrades, a proprietary health care rating company, publishes hospital-specific rates of "major complications" after cesarean and vaginal singleton deliveries, and after "patient choice" cesarean deliveries.[6] Other vendors, such as Solucient[7] and HealthShare,[8] offer similar products. The validity of all of these approaches depends upon the accuracy of the available data.

Multiple previous studies have evaluated the accuracy of complication codes in Medicare claims data,[9,10,11,12] the Veterans Health Administration's Patient Treatment File,[13,14,15,16] California's Patient Discharge Data Set,[17] and similar data sets from individual hospitals[18,19] or other countries.[20,21,22] In general, these studies have demonstrated substantial disagreement between ICD-9-CM coded complications and medical records, with great variation in coding across hospitals,[23] leading to concerns that administrative data should not be used to compare provider complication rates.[24] However, these studies have focused almost entirely on medical-surgical patients, and have not evaluated the coding of obstetric records.

The current study was conducted to validate the ICD-9-CM coding of obstetric complications in the nation's largest statewide patient discharge data program. We examined the sensitivity and positive predictive value (PPV) of maternal hospital discharge abstracts, using the complete inpatient medical record as the gold standard. We hypothesized that clearly defined, clinically meaningful complications that require physician intervention, such as perineal lacerations and cesarean-related injuries, would be accurately coded in hospital discharge data, whereas more ambiguous complications such as postpartum hemorrhage and endometritis would be poorly coded.

METHODS

The general methods for this study are described in Section 2. In this section, we describe only those aspects of the methods that were specifically related to evaluating the coding of perineal lacerations and other obstetric complications.

We searched the MEDLINE database from 1985 through 2000 to identify clinical trials and case series reporting maternal outcomes of delivery. Additional papers were identified by a clinical advisory panel (which included four obstetricians and/or perinatologists, two family physicians, one obstetric nurse specialist, and one health information professional) and by reviewing reference lists in obstetrics texts and meta-analyses. We excluded papers without abstracts or in languages other than English, studies from developing countries, and studies limited to patients

with unusual procedures or risk factors. We then reviewed abstracts to locate studies with at least 250 patients that reported on postpartum maternal complications and/or readmissions. After discussing these findings with our advisory panel, we developed a comprehensive list of maternal complications, which we mapped to ICD-9-CM using appropriate references[25] with the assistance of two coding professionals. We also tested ICD-9-CM definitions of maternal complications from JCAHO,[1] AHRQ,[4] HealthGrades,[6] and Solucient.[7]

After selecting a stratified random sample of vaginal and cesarean deliveries from acute care hospitals in California, as described in Section 2, we asked each participating hospital to photocopy each sampled record. Each record was reviewed by one of four experienced Accredited Record Technicians or Certified Coding Specialists, who recoded the ICD-9-CM diagnosis and procedure codes, as well as maternal demographic and prenatal data, blinded to the original discharge abstract. A regional coding authority tested these individuals before they were hired, trained and supervised them, and verified at least 10% of each abstractor's records to ensure at least 95% accuracy. Discrepancies were resolved through collective review of appropriate coding references. Because JCAHO and AHRQ have endorsed perineal laceration rates as a quality indicator, differences between hospital-reported and reabstracted data on this outcome were carefully evaluated by two authors (JMK, PSR).

In this Section, we evaluate whether specific complications can be accurately ascertained from the ICD-9-CM codes on hospital discharge abstracts. Accuracy was measured in terms of sensitivity and PPV, using our recoding of hospital charts as the gold standard. We defined sensitivity as the proportion of patients with a complication identified through recoding for whom the same complication was reported on the hospital's original discharge abstract. We defined PPV as the proportion of patients with a complication reported on the hospital's original discharge abstract for whom the same complication was independently found through recoding. We do not report specificity, or the percentage of patients without a complication (according to recoding) who were correctly reported as not having it, because this parameter was never below 97%, and nearly always exceeded 99%. We estimated confidence intervals using the SVYTAB procedure in STATA 7.0, which takes into account both oversampling of cesarean deliveries and clustering of observations within hospitals. These confidence intervals are reported elsewhere.[26]

The study protocol was approved by the appropriate committees at the University of California, Davis and the California Health and Human Services Agency.

RESULTS

Tables 1 and 3 show the validated frequency, sensitivity, and PPV of ICD-9-CM coded complications of obstetric care on California hospital discharge abstracts. To maximize sensitivity, we defined many of these complications broadly using both diagnosis and procedure codes. Unweighted sensitivities and PPVs permit computation of the number of false positive and false negative cases. Weighted values permit extrapolation to the entire population of women who were delivered at acute non-federal hospitals with active obstetric services in California during the study period. However, weighted estimates are less stable because women who were readmitted had a higher sampling probability than women who were not. Therefore, the weighted estimate should be interpreted cautiously when it differs greatly from the corres-

ponding unweighted estimate. Unless otherwise stated, weighted estimates are cited below.

Table 1 shows that both third and fourth degree perineal lacerations were reported accurately, with sensitivities exceeding 90% and PPVs exceeding 65% (or 85% unweighted). A thorough review of 9 of the 12 discrepant cases by the lead nurse abstractor and first author confirmed 4 of the 5 false negatives, but 3 of the 4 apparent false positives were determined to be true positives. Unfortunately, records for 3 of the discrepant cases could not be retrieved when this review was undertaken. Reallocation of the 3 cases mislabeled as false positives would increase the unweighted PPVs of third, fourth, and either third or fourth degree lacerations to 93%, 100%, and 95%, respectively. Repair of a third or fourth degree laceration was also accurately reported, but adding this code to the definition of third or fourth degree laceration only affected the allocation of 3 cases. Other intrapartum surgical complications were less frequent than perineal lacerations, so the corresponding estimates in Table 1 are less reliable. Only two of the four cases of urinary tract injury during cesarean delivery were properly reported. Pelvic hematomas and nonspecific accidental injuries were poorly reported, with sensitivities of 17% and 41%, and PPVs of 68% and 99.7%, respectively.

Table 2 shows the highest reported degree of perineal laceration compared with the highest degree found on reabstraction. The shaded cells, representing agreement between the two data sources, include 91% of vaginal deliveries. In each cell is the number or percentage of women who were confirmed as having the same degree of perineal laceration as reported on their discharge abstract. Reallocation of cases mislabeled as false positives (described above) would increase the unweighted percentage of correctly reported third and fourth degree lacerations from 85% and 94% to 93% and 100%, respectively. Of the 72 disagreements in the table, 36 related to whether the patient suffered no injury or a first degree laceration, and 11 related to whether the patient suffered a first or second degree laceration.

Table 3 describes hospitals' reporting of postpartum complications. Although most of these complications were reported with at least 80% PPV, all suffered from substantial underreporting, with sensitivities of 70% or less. At the extreme, none of four thromboembolic complications was reported on the hospital discharge abstract.

Table 4 summarizes the estimated performance of several complex algorithms for identifying obstetric complications, including AHRQ's Patient Safety Indicators (PSIs),[2] experimental indicators, and legacy indicators from the Healthcare Cost and Utilization Project (still used by HealthShare);[8] HealthGrades' measures of major complications after vaginal and cesarean delivery,[6] and Solucient's measures of vaginal and cesarean complications.[7] To avoid discarding informative data, we applied these algorithms without vendor-recommended denominator exclusions (e.g., HealthGrades focuses on single live-born deliveries). With the exception of AHRQ's PSIs, these algorithms suffer from similar underreporting as the complications we defined in Table 3.

DISCUSSION

The current study represents a comprehensive analysis of the validity of obstetric complication codes on 1,611 discharge abstracts from 52 California hospitals. We found very high levels of

agreement (i.e., sensitivities and PPVs of at least 85%) for third and fourth degree perineal lacerations. Agreement was very high regardless whether we used both diagnosis and procedure codes, or diagnosis codes alone, to ascertain these lacerations. The quality of coding for other intrapartum injuries varied, but these findings are harder to interpret because of small numbers. Most postpartum complications, and both public-sector and commercial algorithms based on complication codes, were subject to substantial underreporting, but relatively little overreporting.

These findings are not surprising, because hospitals are only required to report "conditions that affect patient care in terms of requiring: clinical evaluation; or therapeutic treatment; or diagnostic procedures; or extended length of hospital stay; or increased nursing care and/or monitoring."[27] These criteria encompass any complication that requires surgical correction, such as a third or fourth degree perineal laceration or an iatrogenic injury to the cervix or urinary tract. However, these criteria exclude most pelvic hematomas, many superficial wound infections, and many cases of atelectasis and similar complications.

Health information professionals are further instructed "never (to) code a diagnosis as a complication unless it is stated as such and documented in the medical record by the attending physician."[28] They are warned not to "reach into the medical record to code other conditions for the sake of coding...(if in doubt) contact the physician."[29] The emphasis in these guidelines is for coders to rely upon physician documentation, which may be difficult to interpret, incomplete, or even misleading. In the absence of specific clinical criteria for such codes as "complications of the administration of anesthetic or other sedation in labor and delivery" (668.xx), coders at different hospitals may apply these codes very differently. Even for some straightforward codes, such as accidental puncture or laceration (998.2), ascertainment may vary because coders are instructed that "when a tear is documented in the operative report...the surgeon should be queried as to whether (it) was an incidental occurrence inherent in the surgical procedure or whether the tear should be considered...a complication."[30] Anecdotal evidence from coders indicates that such queries rarely occur, are often unanswered, and may voke medicolegal concerns.

Our results are generally consistent with prior research on the Complications Screening Program (CSP), a complex algorithm for using administrative data to screen for potential complications, which demonstrated that 31% of surgical patients with an ICD-9-CM coded complication did not have supporting clinical evidence, and 19% did not even have a supporting physician note.[10] However, this study did not include obstetric patients. To our knowledge, none of the organizations promoting algorithms for ascertaining obstetric complications based on ICD-9-CM codes has published validation findings, although JCAHO has been involved in such efforts. One study focusing on uterine rupture during trial of labor found that the ICD-9-CM codes traditionally used to identify this complication (665.0x-665.1x)[31,32] had a sensitivity of 64% or less, and a PPV of 51%.[33]

The most important limitation of this research is that our recoded data do not provide an ideal gold standard. However, we selected coders with experience in obstetrics, trained them thoroughly using specific written guidelines, monitored them carefully, and gave them unlimited time to code each record. When cases with discrepant data were thoroughly reviewed by the lead nurse abstractor and the first author, nearly every false negative was confirmed but some false positives were classified as recoding errors. Second, because this study was designed to validate

a published report on risk-adjusted outcomes, our medical records came from 1992 and 1993. ICD-9-CM coding of obstetric complications may have improved in the past decade. The mean number of diagnoses reported on California hospital discharge abstracts increased from 4.50 in 1997 to 5.46 in 2003,[34] although comparable statistics are not available for obstetric abstracts. Finally, our exclusion of very low-volume hospitals from the sampling frame may limit the generalizability of our findings. Fortunately, these hospitals accounted for fewer than 10% of deliveries performed in California hospitals during the study period.

Despite these limitations, our findings have important implications for current methods of evaluating the quality of obstetric care. The hospital discharge database serves as an excellent source of data on third and fourth degree perineal lacerations, confirming the criterion validity of JCAHO's Core Measure and AHRQ's Patient Safety Indicator on this topic. This new evidence complements evidence from prior studies that third and fourth degree perineal lacerations have significant long-term effects on anal sphincter function,[35] and are strongly and consistently associated with the use of episiotomy.[36,37,38] Taken together, these studies suggest that it is quite appropriate for hospitals, accrediting organizations, purchasers of hospital care, and public agencies to use administrative data to monitor the perineal laceration rate. If this rate is significantly higher than expected, providers should reevaluate their use of episiotomy and other components of intrapartum management.

Administrative data seem less useful for monitoring other types of early postpartum complications, such as wound infections, endometritis, and anesthesia-related complications. Our results raise serious questions, for example, about the validity of the complications measures promoted by HealthGrades and Solucient in their online hospital report cards. Similar problems are likely to affect other composite measures of postpartum complications[39,40] and the "labor and delivery" measures proposed by the California Maternal Quality of Care Working Group,[41] which are mostly based on the same ICD-9-CM codes. Given that the mean hospital stay for a vaginal delivery is only 2.1 days,[42] it seems almost fruitless to search discharge abstracts from delivery hospitalizations for these complications, which typically do not become apparent for several days. Readmission records or outpatient claims may be more promising sources of data on potentially preventable postpartum complications.[43]

Table 1. Validity of ICD-9-CM intrapartum complications on California hospital discharge abstracts

Procedure	ICD-9-CM codes	No. (%) women with condition	Sensitivity (%)		Positive Predictive Value (%)	
			Unweighted	Weighted	Unweighted	Weighted
Cesarean delivery	74.0-74.2, 74.4, 74.99	790 (22.3%)	100%	100%	100%	100%
Perineal laceration (vaginal deliveries only)	664.2x (3rd degree)	44 (3.8%)	89%	90%	85%	65%
	664.3x (4th degree)	18 (1.4%)	83%	97%	94%	100%
	664.2x, 664.3x* (3rd or 4th degree)	62 (5.3%)	90%	93%	90%	73%
	664.2x, 664.3x, 75.62 (3rd or 4th degree or repair)	64 (5.3%)	92%	94%	92%	60%
Repair of obstetric laceration of rectum and sphincter ani (vaginal deliveries only)	75.62	54 (4.8%)	83%	51%	85%	41%
Injury, laceration, or repair of urinary tract (cesarean only)	56.82, 56.89, 57.81, 57.93, 75.61, 665.5x	4 (0.1%)	50%	95%	100%	100%
Injury, laceration, or repair of uterus, cervix, high vaginal wall	65.71 (cesarean only), 66.71 (cesarean only), 67.61, 69.41, 70.71, 75.5x, 75.93-75.94, 665.1x-665.4x	26 (0.8%)	39%	41%	91%	99.7%
Pelvic hematoma and other intrapartum injuries	665.7x-665.9x, 998.2, E870.0, E870.8, 54.1x (excluding nonobstetric indications)†	17 (0.3%)	18%	17%	38%	68%

* This definition is functionally equivalent to that of JCAHO Core Measure PR-3, "third or fourth degree laceration," and to that of the AHRQ Experimental Indicator entitled "third or fourth degree obstetric lacerations."

† Nonobstetric indications include appendicitis (540.x-542), diverticulitis (562.11-562.13), abscess of intestine (569.5), and acute cholecystitis (575.0, 575.4).

Table 2. Confirmation of highest degree of perineal laceration on California hospital discharge abstracts

Degree of laceration reported by hospital		**Degree of laceration identified by reabstraction**					
		None	**1**	**2**	**3**	**4**	**Missing or incomplete**
None	Number	485*	8	5	4	1	1
	%	96.2%*	1.6%	1.0%	0.8%	0.2%	0.2%
1	Number	28	121*	6	0	0	4
	%	18%	76%*	4%	0%	0%	3%
2	Number	6	5	84*	1	0	0
	%	6%	5%	88%*	1%	0%	0%
3	Number	5	0	0	39*	2	0
	%	11%	0%	0%	85%*	4%	0%
4	Number	1	0	0	0	15*	0
	%	6%	0%	0%	0%	94%*	0%
Total		525	134	95	44	18	5

* In each shaded cell is the number or percentage of women who were confirmed as having the same degree of perineal laceration as reported on their hospital discharge abstract.

Table 3. Validity of ICD-9-CM postpartum complications on California hospital discharge abstracts

Condition	ICD-9-CM codes	No. (%) women with condition	Sensitivity (%)		Positive Predictive Value (%)	
			Unweighted	Weighted	Unweighted	Weighted
Urinary tract infection	646.62, 646.64, 996.64	13 (0.2%)	38%	20%	63%	41%
Wound infection, disruption, or dehiscence	54.0 (cesarean only),[*] 54.3 (cesarean only),* 54.61 (cesarean only), 86.04,* 86.22,* 86.28,* 040.0 (cesarean only), 674.1x (cesarean only), 674.2x (vaginal only), 674.3x[†]	40 (2.0%)	50%	68%	87%	98%
Endometritis	615.0, 615.9, 670.0x	125 (5.5%)	39%	46%	86%	98%
Any of the above infectious complications		**172 (7.6%)**	**42%**	**51%**	**87%**	**98%**
Spinal anesthesia complications	03.95, 324.1, 349.0	16 (1.1%)	50%	24%	89%	97%
Pulmonary complications	507.0, 514, 518.0, 518.4-518.5, 518.81, 518.82,[‡] 668.0x, 799.0-799.1, 997.3	21 (0.8%)	38%	1%	62%	13%
Other anesthetic complications	668.1x-668.9x, 995.4, E938.x, E945.2	21 (1.2%)	19%	18%	67%	83%
Any of the above anesthesia-related complications		**48 (2.3%)**	**42%**	**19%**	**80%**	**82%**
Retained products, delayed postpartum hemorrhage	69.02, 69.52, 75.4,[*] 75.7,[*] 666.0x, 666.2x, 667.xx	49 (1.5%)	67%	62%	80%	84%
Thromboembolic complications	415.1, 451.1x,[§] 451.2,[§] 451.8x,[§] 453.8,[§] 671.4x, 673.1x-673.8x, 997.2, 999.2	4 (0.2%)	0%	0%		

[*] These procedure codes only qualify if the procedure date was after the date of delivery. All other procedure codes qualify if the procedure date was on or after the date of delivery. Manual removal of retained placenta (75.4) on the date of delivery qualifies only for vaginal deliveries.
[†] Additional codes were tested but did not affect ascertainment of this type of complication: 998.3x, 998.5x, 86.01, 86.59.
[‡] These codes were correct at the time that these deliveries were performed. The correct definition of this condition would now include additional codes.
[§] These diagnosis codes only qualify if not accompanied by a diagnosis of antepartum deep phlebothrombosis (671.3x).

Table 4. Validity of currently available and proposed composite measures of ICD-9-CM obstetric complications on California hospital discharge abstracts

Composite Measure	ICD-9-CM codes	No. (%) women with condition	Sensitivity (%)		Positive Predictive Value (%)	
			Unweighted	Weighted	Unweighted	Weighted
AHRQ PSI: Obstetric trauma – vaginal with instrument	664.3x, 665.3x-665.5x, 75.5x, 75.61, 75.62	14 (1.7%)*	79%	68%	85%	100%
AHRQ PSI: Obstetric trauma – vaginal w/out instrument	664.3x, 665.3x-665.5x, 75.5x, 75.61, 75.62	76 (9.3%)†	50%	37%	86%	95%
AHRQ PSI: Obstetric trauma – cesarean section	664.3x, 665.3x-665.5x, 75.5x, 75.61, 75.62	18 (2.3%)‡	11%	5.4%	67%	94%
AHRQ Experimental: Other obstetric complications	668.xx, 669.1x, 669.3x, 669.4x	35 (2.2%)	34%	20%	67%	83%
AHRQ Experimental: Obstetric wound complications, cesarean	674.1x, 674.3x	37 (4.7%)‡	46%	27%	85%	91%
AHRQ Experimental: Obstetric wound complications, vaginal	674.2x, 674.3x, 664.5, 665.7	10 (1.2%)§	90%	98%	100%	100%
AHRQ Experimental: Postpartum UTI	646.62, 646.64	9 (0.6%)	22%	9.4%	100%	100%
Solucient: Vaginal deliveries with complications	Proprietary	163 (20%)§	58%	49%	88%	95%

* The denominator for this percentage includes only women with forceps or vacuum-assisted deliveries (72.0-72.4, 72.51, 72.53, 72.6-72.9 in reabstracted data).
† The denominator for this percentage includes only women with neither forceps nor vacuum-assisted deliveries (according to reabstracted data).
‡ The denominator for this percentage includes only women with cesarean deliveries (according to reabstracted data).
§ The denominator for this percentage includes only women with vaginal deliveries (according to reabstracted data).

Solucient: C-section deliveries with complications	Proprietary	145 (18%)[‡]	54%	53%	90%	90%
Solucient: VBAC with complications	Proprietary	20 (2.4%)[**]	35%	7.7%	70%	97%
AHRQ HCUP legacy: Obstetrical complications[††]	285.1, 664.3x, 666.0x-666.2x, 668.x0, 668.x1, 668.x2, 668.x4, 669.10-669.12, 669.14, 669.3x, 669.4x, 670.0x, 673.00-673.02, 673.04, 674.00-674.02, 674.04, 674.1x, 674.2x, 675.10-675.12, 675.14	196 (12%)	57%	54%	67%	83%
Health Grades: Major complications, vaginal delivery	287.4, 512.x, 518.4, 518.81, 584.5, 584.8, 584.9, 664.21, 664.31, 665.31, 665.41, 665.51, 666.x2, 666.x4, 666.10, 666.20, 666.30, 668.82, 670.02, 674.32, 785.51, 785.59, 996.31, 996.60, 996.62, 997.0x, 997.1, 997.3-997.5, 997.91, 998.0, 998.11, 998.2-998.81, 998.83, 998.89, 998.9, 999.1-999.9	183 (22%)[§]	67%	58%	91%	91%
Health Grades: Major complications, C-section delivery	285.1, 287.4, 512.x, 518.4, 518.81, 584.5, 584.8, 584.9, 648.22, 666.0x, 666.x2, 666.x4, 667.02, 668.82, 669.42, 670.02, 674.12, 674.32, 785.50, 785.51, 785.59, 996.31, 996.60, 996.62, 997.0x, 997.1, 997.3-997.5, 997.91, 998.0, 998.11, 998.2-998.81, 998.83, 998.89, 998.9, 999.1-999.9	163 (21%)[‡]	55%	47%	64%	79%

[**] The denominator for this percentage includes only women with vaginal deliveries after a prior cesarean (654.2x in reabstracted data).
[††] This indicator is currently used by HealthShare, Inc.

REFERENCES

1. Joint Commission on Accreditation of Healthcare Organizations. Specification Manual for National Hospital Quality Measures (2005). Available at: http://www.jcaho.org/pms/core+measures/aligned_manual.htm. Retrieved December 7, 2004.
2. Agency for Healthcare Research and Quality. AHRQuality Indicators. Available at: http://www.qualityindicators.ahrq.gov/. Retrieved December 7, 2004.
3. Romano PS, Geppert JJ, Davies S, Miller MR, Elixhauser A, McDonald KM. A national profile of patient safety in US hospitals based on administrative data. *Health Aff* 2003; 22(2):154-166.
4. *AHRQ Quality Indicators – Guide to Patient Safety Indicators*. Rockville, MD: Agency for Healthcare Research and Quality, 2003. Version 2.1, Revision 3 (January 17, 2005). AHRQ Pub.03-R203. Available at: http://www.qualityindicators.ahrq.gov/downloads/psi/psi_guide_rev3.pdf .
5. Main EK, Bloomfield L, Hunt G, for the Sutter Health First Pregnancy and Delivery Clinical Initiative Committee. Development of a large-scale obstetric quality improvement program that focused on the nulliparous patient at term. *Am J Obstet Gynecol* 2004;190:1747-58.
6. HealthGrades, Inc. Hospital Report Cards Methodology White Paper: Obstetric Services and Women's Health 2004 Analysis (2000-2002 Data). Available at: http://www.healthgrades.com/media/DMS/html/HospitalOBWomensMethodologyWhitePaperwCodes20002002.html. Retrieved December 7, 2004.
7. *Solucient Risk-Adjusted Outcomes Software. Version 5.2. Delivery Models*. Solucient, LLC: Evanston, IL; 2002-2004.
8. HealthShare, Inc. Data Sources. Available at: http://www.selectqualitycare.com/p-con-sources.htm. Retrieved April 18, 2005.
9. Hartz AJ, Kuhn EM. Comparing hospitals that peform coronary artery bypass surgery: The effect of outcome measures and data sources. *Am J Public Health* 1994;84:1609-1614.
10. Lawthers AG, McCarthy EP, Davis RB, Peterson LE, Palmer RH, Iezzoni LI. Identification of in-hospital complications from claims data. Is it valid? *Med Care* 2000;38(8):785-95.
11. Rosen AK, Geraci JM, Ash AS, McNeff KJ, Moskowitz MA. *Predicting postoperative adverse events of common surgical procedures in the Medicare population.* Prepared for Health Care Financing Administration. Springfield, VA: National Technical Information Service; 1991.
12. Keeler EB, Kahn KL, Bentow SS. *Assessing quality of care for hospitalized Medicare patients with hip fracture using coded diagnoses from the Medicare Provider Analysis and Review files*. Prepared for the Health Care Financing Administration. National Technical Information Service: Springfield, VA; 1992.
13. Geraci JM, Ashton CM, Kuykendall DH, Johnson ML, Wu L. International classification of Diseases, 9th Revision, Clinical Modifications Codes in discharge abstracts are poor measures of complication occurrence in medical inpatients. *Med Care* 1997;35:589-602.
14. Berlowitz DR, Brand HK, Perkins C. Geriatric syndromes as outcome measures of hospital care: can administrative data be used? *J Am Geriatr Soc* 1999;47:692-696.
15. Best WR, Khuri SF, Phelan M, et al. Identifying patient preoperative risk factors and postoperative adverse events in administrative databases: Results from theDepartment of Veterans Affairs National Surgical Quality Improvement Program. *J Am Coll Surg* 2002;194:257–266.

16. Barbour GL. Usefulness of a discharge diagnosis of sepsis in detecting iatrogenic infection and quality of care problems. *Am J Med Qual* 1993;8:2-5.
17. Romano PS, Schembri ME, Rainwater JA. Can administrative data be used to ascertain clinically significant postoperative complications? *Am J Med Qual* 2002;17:145-154.
18. Faciszewski T, Johnson L, Noren C, Smith MD. Administrative databases' complication coding in anterior spinal fusion procedures. What does it mean? *Spine* 1995;20:1783-1788.
19. Massanari RM, Wilkerson K, Streed SA, Hierholzer WJ, Jr. Reliability of reporting nosocomial infections in the discharge abstract and implications for receipt of revenues under prospective reimbursement. *Am J Public Health* 1987;77:561-564.
20. Hawker GA, Coyte PC, Wright JG, Paul JE, Bombardier C. Accuracy of administrative data for assessing outcomes after knee replacement surgery. *J Clin Epidemiol* 1997;50:265-273.
21. Taylor B. Common bile duct injury during laparoscopic cholecystectomy in Ontario: Does ICD-9 coding indicate true incidence? *CMAJ* 1998;158:481-485.
22. Valinsky LJ, Hockey RL, Hobbs MST, et al. Finding bile duct injuries using record linkage: A validated study of complications following cholecystecomy. *J Clin Epidemiol* 1999;52:893-901.
23. Romano PS, Chan BK, Schembri ME, Rainwater JA. Can administrative data be used to compare postoperative complication rates across hospitals? *Med Care* 2002;40:856-867.
24. Geraci JM. The demise of comparative provider complication rates derived from ICD-9-CM code diagnoses. *Med Care* 2002;40:847-850
25. *American Hospital Association's ICD-9-CM Coding Handbook, 2004 Revised Edition*. American Hospital Association: Chicago, IL; 2004.
26. Romano PS, Yasmeen S, Schembri ME, Keyzer JM, Gilbert WM. Coding of perineal lacerations and other complications of obstetric care in hospital discharge data. *Obstet Gynecol* 2005;106:717-725.
27. *Coding Clinic for ICD-9-CM* 1990;7(2):13.
28. *Coding Clinic for ICD-9-CM* 1994;11(5):7.
29. *Coding Clinic for ICD-9-CM* 1984;1(1):12.
30. *Coding Clinic for ICD-9-CM* 1990;7(3):18.
31. Gregory KD, Korst LM, Cane P, Platt LD, Kahn K. Vaginal birth after cesarean and uterine rupture rates in California. *Obstet Gynecol* 1999;94:985-989.
32. Lydon-Rochelle M, Holt VL, Easterling TR, Martin DP. Risk of uterine rupture during labor among women with a prior cesarean delivery. *N Engl J Med* 2001;345:3-8.
33. Use of hospital discharge data to monitor uterine rupture--Massachusetts, 1990-1997. *MMWR Morb Mortal Wkly Rep* 2000;49(12):245-8.
34. California Office of Statewide Health Planning and Development, Healthcare Quality & Analysis Division. Find Data – ICD-9-CM Code Frequencies. Available at: http://www.oshpd.ca.gov/HQAD/PatientLevel/ICD9Codes.htm. Retrieved December 6, 2004.
35. Sultan AH, Kamm MA, Hudson CN, Bartram CI. Third degree obstetric anal sphincter tears: risk factors and outcome of primary repair. *BMJ* 1994;308(6933):887-91.
36. Homsi R, Daikoku NH, Littlejohn J, Wheeless CR Jr. Episiotomy: risks of dehiscence and rectovaginal fistula. *Obstet Gynecol Surv* 1994;49(12):803-8.
37. Woolley RJ. Benefits and risks of episiotomy: a review of the English-language literature since 1980. Part I. *Obstet Gynecol Surv* 1995;50(11):806-20.

38. Myers-Helfgott MG, Helfgott AW. Routine use of episiotomy in modern obstetrics. Should it be performed? *Obstet Gynecol Clin North Am* 1999;26(2):305-25.
39. Koroukian SM. Relative risk of postpartum complications in the Ohio Medicaid population: Vaginal versus cesarean delivery. *Med Care Res Rev* 2004;61(2):203-224.
40. Garcia FAR, Miller HB, Huggins GR, Gordon TA. Effect of academic affiliation and obstetric volume on clinical outcome and cost of childbirth. *Obstet Gynecol* 2001;97:567-576.
41. Gregory KD, Hobel C, Korst KM, Lu M, Reyes C. *A Framework for Developing Maternal Quality of Care Indicators.* Cedars-Sinai Medical Center and UCLA School of Medicine: Los Angeles, CA; 2001.
42. *Care of Women in U.S. Hospitals, 2000*. HCUP Fact Book No. 3. AHRQ Publication No. 02-0044, October 2002. Agency for Healthcare Research and Quality, Rockville, MD. Available at: http://www.ahrq.gov/data/hcup/factbk3/factbk3.htm .
43. Lydon-Rochelle M, Holt VL, Martin DP, Easterling TR. Association between method of delivery and maternal rehospitalization. *JAMA* 2000;283:2411-2416.

Section 4: A Systematic Review of the Accuracy of Key Data Elements in Birth Certificate Databases

SUMMARY

Objective: Our objective was to summarize published literature on the validity of key data elements on birth certificates in the United States and other developed nations.

Methods: Following recommendations developed by the Cochrane Collaboration, we searched MEDLINE and EMBASE for studies evaluating the sensitivity or positive predictive value of birth certificate data, based on an external "gold standard."

Results: We found 484 unique citations and selected 100 papers for review, plus two that were identified through extended search methods. The findings of these studies are presented in tabular format. Studies of parental race/ethnicity and age (or date of birth) have reported excellent validity of birth certificate data. The accuracy of these data on parental education and occupation is lower, but at least 69%. Gravidity and parity, especially nulliparity, are accurately reported. Agreement between birth certificates and other data is moderate for timing of the last menstrual period and length of gestation. Almost all pregnancy-related conditions are poorly reported; prior cesarean is the only consistent exception. Birth certificates tend to exaggerate both the duration of prenatal care and the number of visits. Most complications of labor and delivery are reported with moderate sensitivity (20-75%). Method of delivery is accurately reported, but vaginal birth after cesarean (VBAC) is consistently underreported, with sensitivities of 39-70%. Reporting of forceps or vacuum usage has 60-83% sensitivity. Agreement between birth certificates and other data is high (≥93%) for birthweight, Apgar score, birth order, and infant gender. However, abnormal conditions of the newborn and congenital anomalies are consistently underreported, with sensitivities below 50% for most.

Discussion: The validity of obstetric risk factors and outcomes in birth certificate databases is variable, with excellent accuracy for selected data elements. Researchers and policymakers should be cognizant of potential limitations and sources of bias in using these data.

INTRODUCTION

Birth certificates are widely used for epidemiologic research and surveillance throughout North America and Europe, because they are the only source of uniformly collected data on birth outcomes and associated demographic and medical risk factors.[1] For example, birth certificate data are used to track the prevalence of risk factors and health outcomes in pregnant women and newborns,[2,3,4] to monitor the use of medical interventions, [5,6,7] to evaluate the accessibility and quality of perinatal health care,[8,9,10,11,12] to ascertain selected congenital anomalies,[13] and to sample cases for other epidemiologic studies.[14]

Multiple previous studies have evaluated and summarized the accuracy of specific data elements on birth certificates in the United States and various other countries. The results of these studies are highly relevant to anyone who uses birth certificate data.[15] Inadequate appreciation for the limitations of these data may lead to erroneous conclusions about the etiology of adverse perinatal outcomes, trends in public health, and differences in quality of care. However, we are not aware of any comprehensive published summary of this disparate literature.

The current research was undertaken to collect and synthesize the key results of previous studies evaluating the accuracy of demographic and clinical information on birth certificates, focusing principally on the United States but also including comparable systems in other countries. We believed that this information would be useful in selecting and designing risk factors for OSHPD's report on risk-adjusted outcomes of delivery in California.

METHODS

Following recommendations developed by the Cochrane Collaboration,[16] we performed computer searches of both the MEDLINE and EMBASE bibliographic databases[17] using the OVID interface without date or language restrictions. Our last search, using the most comprehensive set of search terms, was performed on 9 December 2004 with the assistance of a professional medical librarian. These terms included "birth certificate(s)" (title, abstract, other term, subheading) in combination with any of the following: comparative study, reliability, reliable, accuracy, accurate, validity, validation, valid, ascertain (truncated), underreport (truncated), disparit (truncated), complete (truncated), measurement error, assessment, sensitivity, or specificity. These search terms were validated by reviewing articles that were known to the authors, reference lists in those articles, and subsequent citations of those articles.

We reviewed abstracts to identify all studies that evaluated the sensitivity and/or positive predictive value (PPV) of demographic and clinical data elements on birth certificates, based on an external "gold standard." Studies in which birth certificates provided a "gold standard" for validating other data sources (e.g., maternal recall), or in which certain data elements on birth certificates were used to validate other data elements on the same records, were set aside. Due to resource limitations, papers without published abstracts (e.g., letters and editorials) were not reviewed, unless the title suggested that it was relevant to our aims. Papers written in languages other than English, French, Spanish, or German; papers focusing on the completeness of birth registration or the prevalence of missing data; and papers focusing on the accuracy of linkage with other data sets, were also set aside. We read but did not abstract papers based on birth

certificates from before 1980, believing that such data would have less relevance to researchers and policy-makers than more recent data. We also read, but did not abstract, papers that simply compared risk factor prevalence across data sets.

We defined sensitivity as the proportion of patients with a condition, according to the "gold standard" database, for whom that condition was reported on the birth certificate. We defined PPV as the proportion of patients with a condition reported on the birth certificate for whom that condition was confirmed using a "gold standard" database. We estimated these values from the published data when they were not actually reported by the authors. In a few cases, we were only able to report percentage agreement with or without the kappa statistic, which represents a measure of agreement adjusted for chance. Values of less than 0.4, 0.4 to 0.6, 0.6 to 0.8, and 0.8 to 1.0 may be interpreted as poor-to-fair, moderate, substantial, and almost perfect agreement, respectively,[18] although this interpretation should be guided by knowledge of prevalence.[19] For quasi-continuously distributed variables such as birthweight and length of gestation, we report the Pearson correlation coefficient (r) instead. Results were stratified by data element to facilitate comparisons across multiple studies. Because of the heterogeneity in methods, the overlap among different samples, and the relatively small number of studies evaluating each data element, we did not attempt a formal meta-analysis.

RESULTS

Our final search yielded 451 citations from MEDLINE and 208 citations from EMBASE, with a total of 484 unique citations. After reviewing titles and abstracts, we selected 100 papers for complete review. An additional three papers and one abstract were identified through extended search methods,[20] as described above. We reviewed data obtained outside the United States, but these data are omitted from the tables because of substantial international differences in birth certificate formatting and data collection procedures. Only sensitivities and PPVs based on at least 14 cases with the characteristic of interest are shown.

Table 1 shows our findings for parental demographic and social characteristics. Six studies of maternal race/ethnicity and two studies of paternal race/ethnicity reported excellent validity of birth certificate data, except for Native American ethnicity (54% sensitivity in California). Three studies of maternal age or date of birth, and one study of paternal age, also reported at least 94% agreement. Three studies of maternal marital status reported conflicting results, with greater than 95% concordance among Tennessee birth certificates from 1989, and at least 89% PPV in the 1988 National Maternal and Infant Health Survey, but only 68% PPV of "married" status among Medicaid-enrolled women in New Jersey in 1989-92. The PPV of birth certificate data on maternal and paternal education was moderate (range 69-95%), and generally better for parents with a college education than for parents with less formal education. In two validation studies on parental occupation, there was 71-72% agreement with mothers' self-report at the level of 3-digit occupational codes from the US Census Bureau, and 76-77% agreement at the level of broader categories (e.g., managerial/ professional, technical, service, etc.). Agreement for paternal occupation was better than that for maternal occupation in one study, largely due to less confusion about whether the father was employed. The major type of error in reporting both the principal source of payment for prenatal care and the expected principal source of payment for delivery care was underreporting of "uninsured" status (34%-44% sensitivity).

Table 2 shows our findings for the outcomes of prior pregnancies. Previous studies have found excellent agreement between birth certificates and other data sources, exceeding 93% in all but case, for both gravidity and parity. Nulliparity has always been identified with at least 96% accuracy, but there has been more error involving ascertainment of grand multiparity. Validity was moderately high for any prior termination of pregnancy (83% sensitivity, 63% PPV), prior infant death (88% sensitivity, 96% PPV), or prior fetal death (64-84% PPV), in one or two studies of each factor. Agreement on the exact number of prior fetal deaths was somewhat poorer.

Table 3 shows our findings for pregnancy-related risk factors and complications. Agreement between birth certificates and other data sources was moderate for both the date last normal menses began and the obstetric estimate of gestation (60%-100% at ±1 week and 78-94% at ±2 weeks). Systematic differences in these durations between data sources were only observed among low birthweight infants in California and Georgia, for whom birth certificates tended to overestimate gestational age. Agreement was at least 89% for plurality, but lower for maternal weight gain. Almost all conditions on the checklist of pregnancy-related risk factors and complications were poorly reported, with sensitivities as low as 0%. Underreporting was least serious for diabetes (42%-83% sensitivity in five studies), pregnancy-induced hypertension (20%-72% sensitivity in five studies), polyhydramnios or oligohydramnios (17-78% sensitivity in five studies), and previous cesarean delivery (81-94% sensitivity in two studies).

Table 4 shows our findings for maternal behaviors and prenatal care. Four studies showed poor agreement (28%-36% in various settings), whereas one study showed moderate agreement (79%), between birth certificates and other data sources on the exact month in which prenatal care began. However, agreement was substantially better (e.g., 51%-92%) when a one-month difference between data sources was classified as agreement or when months were aggregated into trimesters. Similarly, exact agreement on the number of prenatal visits was only 14-38% in four of the five reported studies, but agreement at ±2 visits was moderate (42%-72%). In five of the six relevant studies, birth certificates appeared to overstate usage of prenatal care. Six studies demonstrated moderate underreporting of tobacco use during pregnancy (53%-89% sensitivity in various settings), whereas four studies demonstrated severe underreporting of alcohol use during pregnancy (18%-34% sensitivity, with one outlier reporting 86%).

Table 5 shows our findings for complications of labor and delivery. Most of these conditions are reported with moderate sensitivity, between 20% and 75% in nearly all studies. The best reported conditions include cephalopelvic disproportion (42-77% sensitivity in four studies), placental abruption (29-67% sensitivity in five studies), and breech or other malpresentation (53-71% sensitivity in three studies). For no other condition did the sensitivity of reporting ever exceed 55%. False positive reports on birth certificates were almost invariably less frequent than false negative reports.

Table 6 shows our findings for obstetric procedures and management of delivery. The method of delivery was generally reported accurately on birth certificates, with at least 79% sensitivity and 88% PPV for both primary and repeat cesarean deliveries. However, vaginal birth after cesarean (VBAC) was consistently underreported, with sensitivities of 39-70% in five of the six published

studies. Overreporting of VBAC was far less frequent, with PPVs exceeding 82% in all but one of these studies. Four studies have evaluated birth certificate reporting of forceps or vacuum usage, showing 60-83% sensitivity. Other obstetric procedures are even more poorly reported on birth certificates, with sensitivities of 43-72% for induction of labor.

Table 7 shows our findings for conditions of the newborn and congenital anomalies. Agreement between birth certificates and other data sources was uniformly high (≥93%) for birthweight, Apgar score at 5 minutes, birth order (among multiple gestations), and infant gender. However, abnormal conditions of the newborn and congenital anomalies have been consistently underreported on birth certificates, with sensitivities below 50% for all but the most catastrophic (e.g., anencephaly) and obvious (e.g., cleft lip) anomalies. Overreporting is less problematic, with reported PPVs for these conditions that generally exceed 50%, and often exceed 75%.

Studies based entirely on data from before 1980 are not summarized in these tables, but their results are generally consistent. Most of these earlier studies documented substantial underreporting of congenital anomalies,[21,22,23,24,25,26] a problem that was thoroughly reviewed in the late 1980's.[27] We found and reviewed eight studies from outside the United States. Barry reported that "Notification of Birth Forms" from six Irish hospitals in 1986 were at least 82% accurate for maternal date of birth and marital status, paternal occupation, parity, prior fetal deaths, prior terminations, birthweight, method of delivery, and method of feeding.[28] The accuracy of gestational age and prenatal care data was poorer but comparable to that reported in the United States. A French study reported minimal overreporting but substantial underreporting of cesarean delivery (61-72% sensitivity), instrumented vaginal delivery (39-58% sensitivity), and prematurity (61% sensitivity) on "health certificates at birth" from three maternity wards.[29] Both the sensitivity (65%) and PPV (78%) of reported prematurity were suboptimal in a validation study from Barcelona.[30] The Danish Medical Birth Register was shown to have a sensitivity of 53%, 66%, 71%, and 54% for placenta previa, abruption, polyhydramnios, and hypertensive disorder, respectively, with kappa values between 0.54 and 0.66.[31] Concordance on gestational age in Denmark was 87% within one week and 96% within two weeks. The Finnish Medical Birth Registry was found to have excellent validity (>90% agreement) for place and time of birth, method of delivery, maternal marital status and occupation, smoking, date of last menstruation (±2 days), plurality, parity, gravidity, prior fetal death, infant gender, birthweight, and length (±1 cm), Apgar score (1 minute), and timing of first prenatal visit. Just as in the US, accuracy was lower for length of gestation (86% agreement), number of prenatal visits (62%), previous cesarean (68%), and tocolysis (44%).[32,33] A Taiwanese team reported "fault data rates" below 10% for nearly every data element on birth certificates from ten hospitals, but sensitivities and PPVs cannot be estimated from their publication.[34] In a more detailed analysis from one Taiwanese hospital, concordance was reported as 99%, 94%, 88%, and 95% for infant gender, birth order, gestational age, and birthweight, respectively.[35] Sensitivity of reporting was 97% for low birthweight and 93% for preterm delivery.

DISCUSSION

In this systematic review of the accuracy of information reported on birth certificates, we found that accuracy varies widely across data elements, but is generally consistent across American states. Specifically, maternal demographic and social factors such as date of birth, education,

gravidity, parity, and sources of payment are accurately reported. Method of delivery, Apgar score at 5 minutes, plurality, and infant gender are also accurately reported. Agreement between birth certificates and other data sources is generally weaker for gestational age, maternal behaviors, and use of prenatal care, but is substantially enhanced by relaxing the definition of agreement (e.g., trimester versus month of first prenatal care visit, ±2 prenatal visits, ±2 weeks of gestation). Most specific complications of pregnancy, labor and delivery, and nearly all abnormal conditions of the newborn and congenital anomalies, are seriously underreported on birth certificates. However, most of these conditions are reported with high PPVs, indicating that overreporting is a minor issue.

This review is subject to several limitations. First, we may have failed to identify all relevant references. However, we followed recommendations from the Cochrane Collaborative to maximize the sensitivity of our search, accessing multiple databases and reviewing the reference lists and subsequent citations of relevant papers. The most significant source of missing data may be the "gray" or unpublished literature, which we were unable to search systematically. Although we reviewed the web pages of the National Center for Health Statistics, we were unable to review the web pages of every state health department and national health ministry. Given the general consistency of published data, it seems unlikely that including more unpublished data would significantly alter our conclusions.

Other limitations of our review reflect the limitations of the empirical research upon which it is based. Several of these studies were secondary analyses conducted with available data that were collected for another research project;[36] hence, their results may have limited generalizability to statewide or national birth registries. Other studies had samples that were too small or limited to a single hospital. Small samples are especially problematic in evaluating the accuracy with which rare outcomes, such as congenital anomalies, are reported. Some studies were based on data from before 1989, when the US Standard Certificate of Live Birth was extensively revised. However, more recent studies have generally yielded similar results (Tables 1-7). Finally, in some cases, the comparison data used as the "gold standard" were never independently validated, raising questions about whether that "gold standard" is really more accurate than the birth certificate. For example, some PPVs in Tables 3, 5, 6, and 7 may be biased downward by underascertainment in the ostensible "gold standard" dataset.

Our findings have several implications for OSHPD's maternal outcomes studies. First, birth certificates alone cannot be relied upon to generate valid data on the prevalence or distribution of pregnancy-related risk factors, complications of labor and delivery, minor obstetric procedures, conditions of the newborn, or congenital anomalies. Almost all of these conditions are substantially underreported on birth certificates. If such underreporting or misclassification is nondifferential with respect to other factors of interest, then risk estimates based on birth certificates may be biased toward the null. However, if underreporting varies across settings of care, which also vary in the prevalence of other factors of interest, then the direction of bias is unpredictable. At least one author found that adverse outcomes of prior pregnancies (e.g., preterm birth, low birthweight) were more likely to be reported if the index pregnancy had a similar outcome.[1] Given such evidence of bias, epidemiologists should be especially cautious in their use of these variables.

By contrast, birth certificates can be relied upon for information on parental demographic characteristics, parity and gravidity, plurality, cesarean delivery, birthweight, and Apgar scores. Birth certificate data on the last menstrual period, the estimated length of gestation, and the initiation of prenatal care are acceptable for most purposes, but only if these variables are categorized or otherwise analyzed to minimize the impact of small reporting errors. Similarly, analyses of parental education should be based on specific, socially meaningful cutoffs (e.g., 12 years) to minimize bias. Given the potential utility of these variables for epidemiologic research and surveillance, states that do not collect all of them should be encouraged to do so. For example, California's birth certificate does not include Apgar scores, ostensibly due to concerns about patient privacy (R. Williams, personal communication, 8/28/02).

Third, there is an urgent need to improve how birth certificate data are collected and recorded. Surveys in two states have demonstrated substantial variability in how hospitals accomplish these tasks.[37,38] Data collectors should rely upon a preferred primary source for each element. For example, hospital prenatal records systematically undercount the number of prenatal visits, if they are sent before the patient's last visit, whereas maternal self-report may exaggerate the percentage of women who receive early prenatal care.[37] The accuracy of information about the last menstrual period, prepregnancy weight, and congenital malformations may also vary across data sources.[37] For data that are best collected by asking mothers, specific question formats should be evaluated and implemented.[39] Brief, standardized definitions of unclear concepts may be helpful.[40] In both cross-sectional and longitudinal studies, hospitals achieve better reporting of congenital malformations when they use worksheets that are reviewed and cross-checked with medical records, instead of relying on physicians to list malformations directly on the birth record.[37] Registered nurses may have an important role, as one qualitative study at five hospitals found serious deficiencies in the training and knowledge of non-clinical staff members who were responsible for data collection.[40] A controlled trial from Finland showed that using check-box forms instead of open-ended forms increased the proportion of obstetric procedures that were correctly reported.[41] Similarly, reporting of preeclampsia, placental abruption, breech presentation, and forceps in Washington dramatically improved after check-box forms were implemented, but reporting remained poor for congenital malformations, for which open-ended reporting was retained.[42] Anecdotally, computerized databases may facilitate timely entry and validation of birth certificate data.[43,35]

Finally, these findings provide strong justification for using linked databases that include both maternal discharge abstracts and newborn birth certificates. By searching for conditions of interest in both data sources, users may overcome the problem of underreporting on birth certificates. For example, using the linked database instead of birth certificates alone to ascertain perinatal risk factors in California in 2000-2002, the estimated prevalence of placental abruption, chronic hypertension, eclampsia, anemia, and maternal cardiac disease increased by more than 200%, and the estimated prevalence of diabetes, preterm labor, premature rupture of membranes, preeclampsia, prolonged labor, maternal seizures and renal disease increased by 100-199%. Smaller increases in prevalence were noted for breech presentation (52%), induction of labor (70%), placenta previa (71%), and prolonged rupture of membranes (56%). Assuming that these increases are largely due to fewer false negatives, the sensitivity of the linked database may far exceed that of birth certificate data for several conditions in Tables 3, 5, and 6.

REFERENCES

1. Adams M. Validity of birth certificate data for the outcome of the previous pregnancy, Georgia, 1980-1995. *American Journal of Epidemiology* 2001; 154:883-888.
2. Arias E, MacDorman MF, Strobino DM, Guyer B. Annual summary of vital statistics--2002. *Pediatrics* 2003; 112(6 Pt 1):1215-1230.
3. Savitz DA, Zhang J. Pregnancy-induced hypertension in North Carolina, 1988 and 1989. *American Journal of Public Health* 1992; 82:675-679.
4. Martin JA, Hamilton BE, Sutton PD, Ventura SJ, Menacker F, Munson ML. Births: Final data for 2002. *National Vital StatisticsReports* 2003; 52(10).
5. Menacker F, Curtin SC. Trends in cesarean birth and vaginal birth after previous cesarean, 1991–99. *National Vital StatisticsReports* 2001; 49(13).
6. Curtin SC, Park MM. Trends in the attendant, place, and timing of births, and in the use of obstetric interventions: United States, 1989–97. *National Vital StatisticsReports* 1999; 47(27).
7. MacDorman MF, Mathews TJ, Martin JA, Malloy MH. Trends and characteristics of induced labour in the United States, 1989-98. *Paediatric and Perinatal Epidemiology* 2002; 16:263-273.
8. Gould JB. Vital records for quality improvement. *Pediatrics* 1999; 103:278-290.
9. Katz SJ, Armstrong RW, LoGerfo JP. The adequacy of prenatal care and incidence of low birthweight among the poor in Washington State and British Columbia. *American Journal of Public Health* 1994; 84:986-991.
10. DiGiuseppe DL, Aron DC, Payne SM, Snow RJ, Dierker L, Rosenthal GE. Risk adjusting cesarean delivery rates: a comparison of hospital profiles based on medical record and birth certificate data. *Health Services Research* 2001; 36:959-977.
11. Moore P, Hepworth JT. Use of perinatal and infant health services by Mexican-American Medicaid enrollees. *JAMA* 1994; 272:297-304.
12. Schulman ED, Sheriff DJ, Momany ET. Primary care case management and birth outcomes in the Iowa Medicaid program. *American Journal of Public Health* 1997; 87:80-84.
13. Mathews TJ, Honein MA, Erickson JD. Spina bifida and anencephaly prevalence—United States, 1991-2001. *Morbidity and Mortality Weekly Report Recommendations and Reports* 2002; 51(RR-13).
14. Phares TM, Morrow B, Lansky A, Barfield WD, Prince CB, Marchi KS *et al.* Surveillance for disparities in maternal health-related behaviors--selected states, Pregnancy Risk Assessment Monitoring System (PRAMS), 2000-2001. *Morbidity and Mortality Weekly Report Recommendations and Reports* 2004; Surveillance Summaries 53:1-13.
15. Kirby RS. The quality of vital perinatal statistics data, with special reference to prenatal care. *Paediatric and Perinatal Epidemiology* 1997; 11:122-128.
16. Alderson P, Green S, Higgins JPT, editors. Formulating the problem. Cochrane Reviewers' Handbook 4.2.2; Section 4. In: The Cochrane Library. Chichester, UK: John Wiley & Sons, 2004 (Issue 1).
17. Sampson M, Barrowman NJ, Moher D, Klassen TP, Pham B, Platt R, et al. Should meta-analysts search Embase in addition to Medline? *Journal of Clinical Epidemiology* 2003; 56:943-955.
18. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33:159-174.
19. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol* 1993; 46:423-429.

20. Savoie I, Helmer D, Green CJ, Kazanjian A. Beyond Medline: reducing bias through extended systematic review search. *International Journal Technology Assessment in Health Care* 2003; 19:168-178.
21. Mackeprang M, Hay S, Lunde AS. Completeness and accuracy of reporting of malformations on birth certificates. *Health Services and Mental Health Administration Heath Reports* 1972; 87:43-49.
22. Venters M, Schacht L, Bensel RT. Reporting of Down's syndrome from birth certificate data in the State of Minnesota. *American Journal of Public Health* 1976; 66:1099-1100.
23. Gregg JB, Stanage WF, Johnson W. Birth certificate data: how reliable? *South Dakota Journal of Medicine* 1984; 37:21-22.
24. Hook EB, Albright SG, Cross PK. Use of Bernoulli census and log-linear methods for estimating the prevalence of spina bifida in livebirths and the completeness of vital record reports in New York State. *American Journal of Epidemiology* 1980; 112:750-758.
25. Green HG, Nelson CJ, Gaylor DW, Holson JF. Accuracy of birth certificate data for detecting facial cleft defects in Arkansas children. *Cleft Palate Journal* 1979; 16:167-170.
26. Huether CA, Gummere GR, Hook EB, Dignan PS, Volodkevich H, Barg M *et al.* Down's syndrome: percentage reporting on birth certificates and single year maternal age risk rates for Ohio 1970-79: comparison with upstate New York data. *American Journal of Public Health* 1981; 71:1367-1372.
27. Holtzman NA, Khoury MJ. Monitoring for congenital malformations. *Annual Review of Public Health* 1986; 7:237-266.
28. Barry J. An evaluation of the notification of birth form. *Irish Journal of Medical Science* 1989; 158:102-104.
29. Germain J-M, Czernichow P, Josset V, Torre J-P, Marpeau L, Le Meur H *et al.* [Quality of data acceptable for perinatal epidemiology surveillance: Assessment of the health certificate at birth and the national obstetrics medical file. Study in three Seine-Maritime maternities]. *Journal de Gynecologie, Obstetrique et Biologie de la Reproduction* 1998; 27:384-388.
30. Font M, Pasarin MI, Ricart M, Martos D. [Exactness of the birth registry of Barcelona regarding birth weight and weeks of gestation]. *Gaceta Sanitaria* 2000; 14:386-390.
31. Kristensen J, Langhoff-Roos J, Skovgaard LT, Kristensen FB. Validation of the Danish birth registration. *Journal of Clinical Epidemiology* 1996; 49:893-897.
32. Teperi J. Multi method approach to the assessment of data quality in the Finnish Medical Birth Registry. *Journal of Epidemiology & Community Health* 1993; 47:242-247.
33. Gissler M, Teperi J, Hemminki E, Merilainen J. Data quality after restructuring a national medical registry. *Scandinavian Journal of Social Medicine* 1995; 23:75-80.
34. Lu JH, Lin FM, Shen WY, Chen SJ, Hwang BT, Wu SI *et al.* Data quality of a computerized medical birth registry. *Medical Informatics* 1994; 19:323-330.
35. Lin CM, Lee PC, Teng SW, Lu TH, Mao IF, Li CY. Validation of the Taiwan Birth Registry using obstetric records. *Journal of the Formosan Medical Association* 2004; 103:297-301.
36. Kirby RS. Invited commentary: using vital statistics databases for perinatal epidemiology: does the quality go in before the name goes on? *American Journal of Epidemiology* 2001; 154:889-890.
37. Land G, Vaughan W. Birth certificate completion procedures and the accuracy of Missouri birth certificate data. *Journal of the American Medical Record Association* 1984; 55:31-34.

38. Smulian JC, Ananth CV, Hanley ML, Knuppel RA, Donlen J, Kruse L. New Jersey's electronic birth certificate program: variations in data sources. *American Journal of Public Health* 2001; 91:814-816.
39. Kharrazi M, Epstein D, Hopkins B, Kreutzer R, Doebbert G, Hiatt R *et al.* Evaluation of four maternal smoking questions. *Public Health Reports* 1999; 114:60-70.
40. Northam S, Polancich S, Restrepo E. Birth certificate methods in five hospitals. *Public Health Nursing* 2003; 20:318-327.
41. Teperi J, Makela M, Hemminki E. Controlled trial on medical birth notification design. *Methods of Information in Medicine* 1991; 30:124-126.
42. Frost F, Starzyk P, George S, McLaughlin JF. Birth complication reporting: the effect of birth certificate design. *American Journal of Public Health* 1984; 74:505-506.
43. Costakos DT, Love LA, Kirby RS. The computerized perinatal database: are the data reliable? *American Journal of Perinatology* 1998; 15:453-459.

Table 1. Validity of parental demographic and social characteristics on birth certificates in the United States, 1980-2004

Data Element	Response option or finding	Sensitivity[a]	PPV	Agreement	Setting and sample size (entire study)
Mothers' race/ethnicity	African-American Asian/Pacific Islander White/Caucasian Hispanic Native American	95% 95% 94% 99% 54%	97% 96% 97% 96% 96%		California, 16 of 19 randomly selected hospitals, 1994-95 (N=7,428)[1]
	Exact value			>99%	Tennessee, stratified random sample of live births, 1989 (N=1,016 high-risk cases; N=700 low-risk controls)[2]
	African-American Hispanic	96% 91%	96% 94%		New Jersey, HealthStart program for high-risk Medicaid-eligible women, 1989-92 (N=46,437)[3]
	Exact value			95%, $\kappa = 0.87$	NE Ohio, random sample of deliveries at 20 hospitals, 1993-95 (N=33,616)[4]
	African-American White Hispanic		98% 98% 98%		National Maternal and Infant Health Survey, stratified random sample of live births, 1988 (N=9,953)[5]
	African-American White		96% 96%		National Natality Survey, stratified random sample of live births, 1980 (N=9,941)[6]
Mother's nativity	Native born (whites) Foreign born (whites)		99.8% 95.8%		National Maternal and Infant Health Survey, stratified random sample of live births, 1988 (N=9,953)[5]
Mother's marital status	Married vs. unmarried			>95%	Tennessee, stratified random sample of live births, 1989 (N=1,016 high-risk cases; N=700 low-risk controls)[2]
	Married	89%	68%		New Jersey, HealthStart program for high-risk Medicaid-eligible women, 1989-92 (N=46,437)[3]

	Married (whites) Married (African-Americans) Unmarried (whites) Unmarried (African-Americans)		97% 92% 89% 95%		National Maternal and Infant Health Survey, stratified random sample of live births, 1988 (N=9,953)[5]
Mother's date of birth/age	Exact value			>95%	Tennessee, stratified random sample of live births, 1989 (N=1,016 high-risk cases; N=700 low-risk controls)[2]
	Age at delivery, years (whites) Age at delivery, years (African-Americans)			98% 97%	National Maternal and Infant Health Survey, stratified random sample of live births, 1988 (N=9,953)[5]
	Age at delivery, years (whites) Age at delivery, years (African-Americans)			99% 98%	National Natality Survey, stratified random sample of live births, 1980 (N=9,941)[6]
Mother's education	Completed years (exact value)			r = 0.86	California, 16 of 19 randomly selected hospitals, 1994-95 (N=10,055)[7]
	Category (whites) Category (African-Americans) ≤11 years (whites) ≤11 years (African-Americans) 12 years (whites) 12 years (African-Americans) 13-15 years (whites) 13-15 years (African-Americans) ≥16 years (whites) ≥16 years (African-Americans)		 84% 80% 86% 78% 83% 80% 93% 86%	86% 80%	National Maternal and Infant Health Survey, stratified random sample of live births, 1988 (N=9,953)[5]

	Completed years (whites) Completed years (African-Americans) 0-8 years (whites) 8-11 years (whites) 8-11 years (African-Americans) 12 years (whites) 12 years (African-Americans) 13-15 years (whites) 13-15 years (African-Americans) ≥16 years (whites) ≥16 years (African-Americans)		 75% 87% 77% 91% 84% 79% 85% 94% 93%	88% 84%	National Natality Survey, stratified random sample of live births, 1980 (N=9,941)[6]
Mother's occupation	Exact value (3-digit census code) General agreement (categories)			72% 76%	New York, cases with anencephaly, spina bifida, cleft lip/palate, and randomly sampled matched controls, 1983-86 (N=1,760) [8]
	Exact value (3-digit census code) Employed (vs. unemployed) General agreement (among employed)	77%	91%	71% 76% 77%	California (Santa Clara), cases with congenital heart disease (N=155) and randomly sampled controls (N=176), 1981-83[9]
Mother's industry	Exact value General agreement			78% 78%	New York, cases with anencephaly, spina bifida, cleft lip/palate, and randomly sampled matched controls, 1983-86 (N=1,760)[8]
Father's race/ethnicity	African-American White Hispanic		98% 98% 96%		National Maternal and Infant Health Survey, stratified random sample of live births, 1988 (N=9,953)[5]
	African-American White		95% 94%		National Natality Survey, stratified random sample of live births, 1980 (N=9,941)[6]
Father's date of birth/age	Age at delivery, years (whites) Age at delivery (African-Americans)			99% 94%	National Natality Survey, stratified random sample of live births, 1980 (N=9,941)[6]

Father's education	Category (whites) Category (African-Americans) ≤11 years (whites) ≤11 years (African-Americans) 12 years (whites) 12 years (African-Americans) 13-15 years (whites) 13-15 years (African-Americans) ≥16 years (whites) ≥16 years (African-Americans)		 83% 72% 84% 78% 75% 69% 93% 80%	84% 75%	National Maternal and Infant Health Survey, stratified random sample of live births, 1988 (N=9,953)[5]
	Category (whites) Category (African-Americans) 0-8 years (whites) 8-11 years (whites) 8-11 years (African-Americans) 12 years (whites) 12 years (African-Americans) 13-15 years (whites) 13-15 years (African-Americans) ≥16 years (whites) ≥16 years (African-Americans)		 75% 84% 78% 87% 79% 77% 72% 95% 83%	86% 78%	National Natality Survey, stratified random sample of live births, 1980 (N=9,941)[6]
Father's occupation	Exact value (3-digit census code) Employed (vs. unemployed) General agreement (among employed)	 96%	 99%	80% 89% 73%	California (Santa Clara), cases with congenital heart disease (N=155) and randomly sampled controls (N=176), 1981-83[9]
Principal source of payment for prenatal care	Private Capitated private MediCal Other Uninsured	97% 84% 96% 75% 34%		$\kappa = 0.90$	California, 16 of 19 randomly selected hospitals, 1994-95 (N=7,428)[1]
Expected principal source of payment for delivery	Private Capitated private MediCal Other Uninsured	98% 85% 96% 83% 44%		$\kappa = 0.92$	California, 16 of 19 randomly selected hospitals, 1994-95 (N=7,428)[10]

[a] All statistics shown are based on a denominator of at least 14 cases.

Table 2. Validity of historical data describing previous pregnancy outcomes on birth certificates in the United States, 1980-2004

Data Element	Response option or finding	Sensitivity[a]	PPV	Agreement	Setting and sample size (entire study)
Parity	Exact value Nulliparity	 99.7%		98%, κ = 0.96 99%, κ = 0.98	Washington, stratified random sample of low-risk deliveries with early prenatal care, 1989-90 (N=1,937)[11]
	Exact value (whites) Exact value (African-Americans) Nulliparity (whites) Nulliparity (African-Americans) 2nd live birth (whites) 2nd live birth (African-Americans) 3rd live birth (whites) 3rd live birth (African-Americans) 4th or higher live birth (whites) 4th or higher live birth (AA)		 98% 96% 84% 77% 81% 73% 77% 70%	89% 82%	National Maternal and Infant Health Survey, stratified random sample of live births, 1988 (N=9,953)[5]
	Exact value (whites) Exact value (African-Americans) Nulliparity (whites) Nulliparity (African-Americans) 2nd live birth (whites) 2nd live birth (African-Americans) 3rd live birth (whites) 3rd live birth (African-Americans) 4th or higher live birth (whites)		 99% 96% 98% 97% 96% 87% 91-92%	98% 94%	National Natality Survey, stratified random sample of live births, 1980 (N=9,941)[6]
	Nulliparity	99%	97%	99%, κ =0.97	NE Ohio, random sample of deliveries at 20 hospitals, 1993-95 (N=33,616)[4]
	Children now living, exact value Children now deceased, exact value			>95% >95%	Tennessee, stratified random sample of live births, 1989 (N=1,016 high-risk cases; N=700 low-risk controls)[2]

	Children now deceased, one or more	88%	96%		Georgia, all singleton births with one prior linked birth certificate, 1989-95 (N=130,806)[12]
Gravidity	Exact value Primigravid	99.4%		93%, κ = 0.83 93%, κ = 0.88	Washington, stratified random sample of low-risk deliveries with early prenatal care, 1989-90 (N=1,937)[11]
	Prior pregnancy (any)	95%	99%	96%, κ = 0.90	NE Ohio, random sample of deliveries at 20 hospitals, 1993-95 (N=33,616)[4]
	Primigravid (whites) Primigravid (African-Americans)		92% 86%		National Natality Survey, stratified random sample of live births, 1980 (N=9,941)[6]
Prior fetal deaths (induced or spontaneous)	Exact value (whites) Exact value (African-Americans) One or more (whites) One or more (African-Americans) One (whites) One (African-Americans) Two (whites) Two (African-Americans) Three or more (whites) Three or more (African-Americans)		 65% 64% 70% 51% 58% 35% 52% 35%	81% 78%	National Maternal and Infant Health Survey, stratified random sample of live births, 1988 (N=9,953)[5]
	Exact value (whites) One or more (whites) One or more (African-Americans) One (whites) One (African-Americans) Two or more (whites)		 84% 73% 75% 59% 85%	89%	National Natality Survey, stratified random sample of live births, 1980 (N=9,941)[6]
Prior termination of pregnancy	Any Exact value <20 weeks, exact value	83%	63%	 r = 0.74 r = 0.73	Single teaching hospital, random sample of live births, 1995 (N=99)[13]

[a] All statistics shown are based on a denominator of at least 14 cases.

Table 3. Validity of pregnancy-related risk factors and complications reported on birth certificates in the United States, 1980-2004

Data Element	Response option or finding	Sensitivity[a]	PPV	Agreement	Setting and sample size (entire study)
Plurality	Single, twin, triplet, etc. (exact value)			>99%	Tennessee, stratified random sample of live births, 1989 (N=1,016 high-risk cases; N=700 low-risk controls)[2]
	Singleton (whites) Singleton (African-Americans) Twin or higher (whites) Twin or higher (African-Americans)		99.8% 99.5% 96.4% 89.0%		National Maternal and Infant Health Survey, stratified random sample of live births, 1988 (N=9,953)[5]
Date last normal menses began	Exact month (high-risk) Exact month (low-risk)			82% 88%	Tennessee, stratified random sample of live births, 1989 (N=1,016 high-risk cases; N=700 low-risk controls)[2]
	±1 week (cerebral palsy cases <1500 g) ±1 week (controls <1500 g) ±1 week (CP cases 1500-2499 g) ±1 week (controls 1500-2499 g) ±1 week (CP cases >2499 g) ±1 week (controls >2499g)			68% 60% 60% 100% 77% 78%	California, singleton children surviving to age 3 with CP (N=172) and random sample of controls (N=472) from 4 counties, 1983-85[14]
	±1 week			91%	Washington, stratified random sample of low-risk deliveries with early prenatal care, 1989-90 (N=1,937)[11]
	Exact week ±1 week ±2 weeks			26% 64% 78%	Georgia, cases with birthweight <1500 g and controls with birthweight >2500 g, 1986-88 (N=1,311)[15]
	Exact date ±1 week			87% 93%	New York, random sample of birth certificates from 4 counties, 1999 (N=400)[16]
Clinical estimate of length of gestation	Exact value (high-risk) Exact value (low-risk) ±1 week (high-risk) ±1 week (low-risk) ±2 weeks (high-risk) ±2 weeks (low-risk)			40% 44% 65% 79% 79% 94%	Tennessee, stratified random sample of live births, 1989 (N=1,016 high-risk cases; N=700 low-risk controls)[2]

	Exact value			r = 0.68 (mean difference <0.1 weeks)	New Jersey, HealthStart program for high-risk Medicaid-eligible women, 1989-92 (N=46,437)[3]
	Preterm vs. term vs. postterm			95%, κ = 0.73	NE Ohio, random sample of deliveries at 20 hospitals, 1993-95 (N=33,616)[4]
	Exact value (whites) Exact value (African-Americans) Preterm (<37 weeks) (whites) Preterm (African-Americans) Full-term (37 or more weeks) (whites) Full-term (African-Americans)		 90% 89% 73% 79%	r = 0.80 r = 0.84	National Natality Survey, stratified random sample of live births with oversampling of LBW, 1980 (N=9,941)[17,6]
Weight gain during pregnancy	Exact value			83%	North Carolina, stratified random sample of 42 hospitals, 1989 (N=395)[18]
	Exact value			r = 0.57 (mean difference 2.2 lbs.)[b]	New Jersey, HealthStart program for high-risk Medicaid-eligible women, 1989-92 (N=46,437)[3]
Pregnancy-related conditions	Previous macrosomia Previous low birthweight	7% 10%	78% 71%		Georgia, all singleton births with one prior linked birth certificate, 1989-95 (N=130,806)[1]
	Overall agreement across all checklist items			59%	North Carolina, stratified random sample of 42 hospitals, 1989 (N=395)[18]
	Poly/oligohydramnios Chronic hypertension Pregnancy-induced hypertension Diabetes	30% 7% 59% 52%		96%, κ = 0.43 98%, κ = 0.12 95%, κ = 0.58 98%, κ = 0.66	Washington, stratified random sample of low-risk deliveries with early prenatal care, 1989-90 (N=1,937)[11]

	Anemia Cardiac disease Acute/chronic lung disease Diabetes mellitus Poly/oligohydramnios Chronic hypertension Pregnancy-induced hypertension Incompetent cervix Previous macrosomia Previous preterm or SGA infant Renal disease Uterine bleeding Eclampsia Hemoglobinopathy Genital herpes Rh sensitization	12% 10% 8% 42% 17% 19% 20% 20% 6% 11% 3% 0.4% 5% 1% 11% 3%	44% 41% 45% 69% 15% 35% 38% 57% 23% 39% 11% 13% 20% 6% 56% 29%		New Jersey, HealthStart program for high-risk Medicaid-eligible women, 1989-92 (N=46,437)[3]
	Anemia Cardiac disease Acute/chronic lung disease Gestational diabetes Poly/oligohydramnios Chronic hypertension Pregnancy-induced hypertension Incompetent cervix Previous macrosomia Previous SGA infant Renal disease Uterine bleeding Eclampsia Genital herpes Previous cesarean	11% 11% 18% 46% 21% 32% 34% 39% 12% 20% 15% 9% 10% 33% 81%	33% 26% 51% 72% 68% 37% 57% 77% 64% 70% 9% 11% 26% 57% 95%	95.3%, $\kappa = 0.14$ 98.7%, $\kappa = 0.15$ 96.8%, $\kappa = 0.25$ 97.1%, $\kappa = 0.55$ 98.0%, $\kappa = 0.32$ 99.0%, $\kappa = 0.34$ 96.3%, $\kappa = 0.40$ 99.6%, $\kappa = 0.52$ 93.9%, $\kappa = 0.19$ 95.0%, $\kappa = 0.29$ 99.5%, $\kappa = 0.11$ 97.7%, $\kappa = 0.09$ 99.4%, $\kappa = 0.14$ 98.9%, $\kappa = 0.42$ 96.5%, $\kappa = 0.85$	NE Ohio, random sample of deliveries at 20 hospitals, 1993-95 (N=33,616)[4]
	Previous cesarean Previous macrosomia	94% 12%	“few FP”		New Jersey, sequential linked birth certificates, 1996-2001 (N=75,516)[19]

	Anemia	67%	36%		New York, random sample of birth certificates from 4 counties, 1999 (N=400)[16]
	Cardiac disease	13%	7%		
	Acute/chronic lung disease	33%/0%	83%/0%		
	Gestational/chronic diabetes	83%/50%	83%/50%		
	Poly/oligohydramnios	78%	93%		
	Chronic hypertension	0%	0%		
	Pregnancy-induced hypertension	72%	72%		
	Incompetent cervix	50%	100%		
	Previous macrosomia	6%	50%		
	Previous preterm infant	3%	8%		
	Previous low birthweight	27%	21%		
	Renal disease	55%	75%		
	Uterine bleeding	33%	75%		
	Genital herpes	67%	92%		
	Other sexually transmitted disease	43%	60%		
	Rh sensitization	100%	17%		
	Preeclampsia	62%	100%		
	Thyroid condition	80%	100%		
	Anemia	15-22%	24-43%		Tennessee, stratified random sample of live births, 1989 (N=1,016 high-risk cases; N=700 low-risk controls)[2]
	Cardiac disease	8-13%	40-100%		
	Acute/chronic lung disease	10-12%	60-100%		
	Diabetes mellitus	65-74%	85%		
	Polyhydramnios	34%	75%		
	Oligohydramnios	30%	73%		
	Chronic hypertension	41%	83%		
	Pregnancy-induced hypertension	43-49%	78-91%		
	Incompetent cervix	52%	85%		
	Previous macrosomia	16-35%	50-64%		
	Previous preterm infant	32-34%	62-75%		
	Previous SGA infant	18%	31%		
	Renal disease	21%	50%		
	Uterine bleeding	12-20%	45-62%		

[a] All statistics shown are based on a denominator of at least 14 cases.
[b] Mean maternal weight gain was understated on birth certificates.

Table 4. Validity of maternal behaviors and prenatal care utilization reported on birth certificates in the United States, 1980-2004

Data Element	Response option or finding	Sensitivity[a]	PPV	Agreement	Setting and sample size (entire study)
Month of pregnancy prenatal care began	Exact value Trimester			79% 92%	North Carolina, stratified random sample of 42 hospitals, 1989 (N=395)[18]
	Exact value (high-risk) Exact value (low-risk) Trimester (high-risk) Trimester (low-risk)			34%[b] 29%[b] 64% 67%	Tennessee, stratified random sample of live births, 1989 (N=1,016 high-risk cases; N=700 low-risk controls)[2]
	Exact value Trimester Care begun in 1st trimester Care begun in 3rd trimester	 95% 23%	 45% 82%	31%[b] 51%	NE Georgia, all women treated by CNMs in public clinic, 1980-88 (N=2,032)[20]
	Exact value ±1 month No prenatal care		 59%	35%[b] 75%	National Natality Survey of birth certificates and physicians, 1980[21]
	Exact value (whites) Exact value (African-Americans) ±1 month (whites) ±1 month (African-Americans)			36%, r = 0.60 [b] 28%, r = 0.55 [b] 77% 64%	National Natality Survey, follow-back of randomly sampled birth certificates, 1980 (N=9,941)[17]
	Care begun in 1st trimester	82%	56%		New Jersey, HealthStart program for high-risk Medicaid-eligible women, 1989-92 (N=46,437)[3]
	Trimester Any prenatal care	 99.6%	 99.6%	80%, κ = 0.51 99%, κ = 0.67	NE Ohio, random sample of deliveries at 20 hospitals, 1993-95 (N=33,616)[4]
Number of prenatal visits	Exact value			82% [c]	North Carolina, stratified random sample of 42 hospitals, 1989 (N=395)[18]

	Exact value (high-risk) Exact value (low-risk) ±1 (high-risk) ±1 (low-risk) ±2 (high-risk) ±2 (low-risk)			32%[c] 20%[c] 55% 41% 68% 54%	Tennessee, stratified random sample of live births, 1989 (N=1,016 high-risk cases; N=700 low-risk controls)[2d]
	Exact value ±1 ±2			14%[c] 36% 53%	NE Georgia, all women treated by CNMs in public clinic, 1980-88 (N=2,032)[20]
	Exact value (whites) Exact value (African-Americans) ±2 (whites) ±2 (African-Americans)			r = 0.45[c] r = 0.47[c] 44% 42%	National Natality Survey, follow-back of randomly sampled birth certificates, 1980 (N=9,941)[17]
	Exact value			r = 0.59 (mean difference 0.2 visits)[c]	New Jersey, HealthStart program for high-risk Medicaid-eligible women, 1989-92 (N=46,437)[3]
	Exact value ±1 ±2			22%, κ = 0.12 53% 72%	Washington, stratified random sample of low-risk deliveries with early prenatal care, 1989-90 (N=1,937)[11]
	≥5 (vs. 1-5 vs. 0)			93%, κ = 0.53	NE Ohio, random sample of deliveries at 20 hospitals, 1993-95 (N=33,616)[4]
	Exact value ±1 ±2			38% 59% 70%	New York, random sample of birth certificates from 4 counties, 1999 (N=400)[16]
Tobacco use	Agreement among women for whom either birth certificate or medical record suggested smoking tobacco			84%	North Carolina, stratified random sample of 42 hospitals, 1989 (N=395)[18]
	Yes	74-78%	95-96%		Tennessee, stratified random sample of live births, 1989 (N=1,016 high-risk cases; N=700 low-risk controls)[2]

	GA AL AK ME SC WV	71% 75% 75% 77% 80% 82%	Assumed 100% Assumed 100% Assumed 100% Assumed 100% Assumed 100% Assumed 100%		AL, AK, GA, ME, SC, WV, stratified random sample of white women after delivery, 1993-95 (N=19,483)[22e]
	Yes	53%	71%		New Jersey, HealthStart program for high-risk Medicaid-eligible women, 1989-92 (N=46,437)[3]
	Yes Cigarettes per day, exact value	64%	80%	r = 0.76 (mean difference 1.2)[f]	Single teaching hospital, random sample of live births, 1995 (N=99)[13]
	Yes	72%	94%	92%, κ = 0.77	NE Ohio, random sample of deliveries at 20 hospitals, 1993-95 (N=33,616)[4]
	Yes	89%	89%		New York, random sample of birth certificates from 4 counties, 1999 (N=400)[16]
Alcohol use	Agreement among women for whom either birth certificate or medical record suggested drinking alcohol			56%	North Carolina, stratified random sample of 42 hospitals, 1989 (N=395)[18]
	Yes	31-34%	75-83%		Tennessee, stratified random sample of live births, 1989 (N=1,016 high-risk cases; N=700 low-risk controls)[2]
	Yes	18%	42%		New Jersey, HealthStart program for high-risk Medicaid-eligible women, 1989-92 (N=46,437)[3]
	Yes	23%	77%	94%, κ = 0.34	NE Ohio, random sample of deliveries at 20 hospitals, 1993-95 (N=33,616)[4]
	Yes	86%	75%		New York, random sample of birth certificates from 4 counties, 1999 (N=400)[16]

[a] All statistics shown are based on a denominator of at least 14 cases.
[b] The birth certificate tended to report earlier initiation of prenatal care than was documented in medical records.
[c] In most cases of disagreement, the number of visits recorded on the birth certificate was higher (or else the mean number of visits was higher on birth certificates).
[d] Failure to obtain prenatal care until the third trimester or obtaining no prenatal care at all was the index with the least percentage of difference (9.7% for the cases and 9.8% for the controls) between the results obtained from birth certificates and those obtained from medical records.
[e] Estimates from this study are derived through a two-sample capture-recapture method, which is based on the arguable assumption that neither data set has any false positive errors.
[f] Birth certificates indicated lower mean consumption than medical records.

Table 5. Validity of labor and delivery complications reported on birth certificates in the United States, 1980-2004

Complication or finding	Sensitivity[a]	PPV	Agreement	Setting and sample size (entire study)
Cephalopelvic disproportion	52%	69%		Washington, 23 volunteer hospitals, 1989 (N=7,536)[23]
Fetal distress	17%	not reported		
Mother febrile (high-risk)	21%	83%		Tennessee, stratified random sample of live births, 1989 (N=1,016 high-risk cases; N=700 low-risk controls)[2]
Mother febrile (low-risk)	9%	42%		
Meconium, moderate-heavy	44-46%	80-84%		
Premature rupture (high-risk)	52%	13%		
Premature rupture (low-risk)	20%	57%		
Prolonged rupture	26-39%	87-100%		
Placental abruption	47%	92%		
Placenta previa	55%	88%		
Other excessive bleeding	5-7%	33-50%		
Precipitous labor (<3 hrs)	31-36%	29-71%		
Prolonged labor (>20 hrs)	23%	56%		
Dysfunctional labor	17%	55%		
Breech presentation	66-71%	91-100%		
Other malpresentation	26%	44%		
Cephalopelvic disproportion	72%	82%		
Prolapsed cord	52%	85%		
Fetal distress	38-39%	68-76%		
Overall agreement across all checklist items			63%	North Carolina, stratified random sample of 42 hospitals, 1989 (N=395)[18]

Mother febrile	10%	21%		New Jersey, HealthStart program for high-risk Medicaid-eligible women, 1989-92 (N=46,437)[3]
Meconium, moderate-heavy	32%	33%		
Premature rupture	20%	35%		
Placental abruption	29%	50%		
Placenta previa	40%	63%		
Other excessive bleeding	6%	18%		
Precipitous labor (<3 hrs)	24%	26%		
Prolonged labor (>20 hrs)	5%	10%		
Dysfunctional labor	17%	30%		
Breech/other malpresentation	53%	63%		
Cephalopelvic disproportion	42%	43%		
Prolapsed cord	21%	19%		
Fetal distress	33%	46%		
Seizure during labor	23%	32%		
Anesthetic complications	0%	0%		
Meconium, moderate-heavy	6%	33%		New York, random sample of birth certificates from 4 counties, 1999 (N=400)[16]
Premature rupture	29%	64%		
Placental abruption	67%	100%		
Precipitous labor (<3 hrs)	33%	75%		
Prolonged labor (>20 hrs)	9%	67%		
Dysfunctional labor	22%	7%		
Meconium, moderate-heavy	39%	76%	90.0%, $\kappa = 0.47$	NE Ohio, random sample of deliveries at 20 hospitals, 1993-95 (N=33,616)[4]
Premature rupture	38%	25%	96.5%, $\kappa = 0.29$	
Placental abruption	52%	67%	99.4%, $\kappa = 0.58$	
Placenta previa	40%	75%	99.5%, $\kappa = 0.52$	
Breech/other malpresentation	65%	86%	98.0%, $\kappa = 0.73$	
Prolapsed cord	24%	39%	99.7%, $\kappa = 0.30$	
Placental abruption	46%		99.5%, $\kappa = 0.58$	Washington, stratified random sample of low-risk deliveries with early prenatal care, 1989-90 (N=1,937)[11]
Placenta previa	49%		99.7%, $\kappa = 0.59$	
Cephalopelvic disproportion	77%		98%, $\kappa = 0.80$	
Fetal distress	22%		84%, $\kappa = 0.29$	

[a] All statistics shown are based on a denominator of at least 14 cases.

Table 6. Validity of obstetric procedures and management of delivery, as reported on birth certificates in the United States, 1980-2004

Data Element	Response option or finding	Sensitivity[a]	PPV	Agreement	Setting and sample size (entire study)
Maternal transport prior to delivery	Yes (high-risk)	57%	97%		Tennessee, stratified random sample of live births, 1989 (N=1,016 high-risk cases; N=700 low-risk controls)[2]
	Yes	18%	63%		New Jersey, HealthStart program for high-risk Medicaid-eligible women, 1989-92 (N=46,437)[3]
	Yes	42%	56%	99.7%, $\kappa = 0.48$	NE Ohio, random sample of deliveries at 20 hospitals, 1993-95 (N=33,616)[4]
Infant transport after delivery	Yes (high-risk)	73%	97%		Tennessee, stratified random sample of live births, 1989 (N=1,016 high-risk cases; N=700 low-risk controls)[2]
	Yes	23%	80%		New Jersey, HealthStart program for high-risk Medicaid-eligible women, 1989-92 (N=46,437)[3]
	Yes	48%	87%	99%, $\kappa = 0.61$	NE Ohio, random sample of deliveries at 20 hospitals, 1993-95 (N=33,616)[4]
Obstetric procedures	Induction of labor	56%	88%		Washington, 23 volunteer hospitals, 1989 (N=7,536)[23]
	Overall agreement across all checklist items			69%	North Carolina, stratified random sample of 42 hospitals, 1989 (N=395)[18]
	Induction of labor Stimulation of labor Tocolysis	43-61% 20-26% 32-37%	63-75% 63-75% 59-89%		Tennessee, stratified random sample of live births, 1989 (N=1,016 high-risk cases; N=700 low-risk controls)[2b]
	Tocolysis	4%	26%		New Jersey, HealthStart program for high-risk Medicaid-eligible women, 1989-92 (N=46,437)[3b]
	Induction or stimulation of labor	72%		86%, $\kappa = 0.68$	Washington, stratified random sample of low-risk deliveries with early prenatal care, 1989-90 (N=1,937)[11b]
Birth attendant	Certified nurse midwife (vs. other)	89%			Michigan, birth logs of 3 nurse mid-wives at one hospital, 1999 (N=97)[24]

Method of delivery	Cesarean Primary cesarean Repeat cesarean Vaginal birth after cesarean Forceps/vacuum	84% 80% 83% 70% 70%	99.8% 99.2% 91.1% 91.0% 95.3%		Washington, 23 volunteer hospitals, 1989 (N=7,536)[23]
	Cesarean	95%	99.5%		Georgia, 3% random sample of singletons from metro Atlanta + 24 other counties, 1986-88 (N=2,423)[25]
	Primary cesarean Repeat cesarean Vaginal birth after cesarean	79% 97% 42%	91%[c] 92%[c] 88%[c]		Georgia, sequential linked birth certificates, 1989-92 (N=106,049)[25]
	Vaginal spontaneous vs. forceps vs. vacuum vs. cesarean			92%[d]	North Carolina, stratified random sample of 42 hospitals, 1989 (N=395)[18]
	Primary cesarean Repeat cesarean Vaginal birth after cesarean Forceps Vacuum	91-93% 79-97% 39-53% 75-83% 71%	96% 93-97% 82-100% 93-94% 83%		Tennessee, stratified random sample of live births, 1989 (N=1,016 high-risk cases; N=700 low-risk controls)[2]
	Primary cesarean Repeat cesarean Vaginal birth after cesarean Forceps Vacuum	81% 80% 47% 60% 60%	88% 88% 55% 41% 50%		New Jersey, HealthStart program for high-risk Medicaid-eligible women, 1989-92 (N=46,437)[3]
	Cesarean Vaginal birth after cesarean	96% 61%	99% 89%	99%, $\kappa = 0.96$ 97%, $\kappa = 0.71$	NE Ohio, random sample of deliveries at 20 hospitals, 1993-95 (N=33,616)[4]
	Primary cesarean Repeat cesarean Vaginal birth after cesarean	98% 100% 100%	100% 100% 100%		New York, random sample of birth certificates from 4 counties, 1999 (N=400)[16]

[a] All statistics shown are based on a denominator of at least 14 cases.
[b] Data on the accuracy of reporting for other obstetric procedures (e.g., amniocentesis, internal electronic fetal monitoring, external electronic fetal monitoring, ultrasound, chorionic villus sampling) are also available here, but are not shown because the US Standard Certificate of Live Birth omits these data elements.
[c] This quantity is estimated from the data reported in the manuscript.
[d] Most discrepancies in this study were due to failure to record use of forceps on the birth certificate.

Table 7. Validity of neonatal characteristics and outcomes reported on birth certificates in the United States, 1980-2004

Data Element	Response option or finding	Sensitivity[a]	PPV	Agreement	Setting and sample size (entire study)
Birthweight	Exact value			100%	North Carolina, stratified random sample of 42 hospitals, 1989 (N=395)[18]
	Exact value (high-risk) Exact value (low-risk)			94% >99%	Tennessee, stratified random sample of live births, 1989 (N=1,016 high-risk cases; N=700 low-risk controls)[2]
	Exact value			r = 0.98 (mean difference 8 g)	Single teaching hospital, random sample of live births, 1995 (N=99)[13]
	Exact value Low birthweight (<2500 g) Very low birthweight (<1500 g)	 91% 85%	 89% 86%	r = 0.94 (mean difference 1 g)	New Jersey, HealthStart program for high-risk Medicaid-eligible women, 1989-92 (N=46,437)[3]
	Exact value ±100 g Very low birthweight (<1500 g)	 90%[b]		93% 98%	Georgia, cases with birthweight <1500 g and controls with birthweight >2500 g, 1986-88 (N=1,311)[15]
	Low birthweight (<2500 g) Very low birthweight (<1500 g)	100% 100%	100% 100%		New York, random sample of birth certificates from 4 counties, 1999 (N=400)[16]
	Exact value (whites) Exact value (African-Americans) 500 g categories (whites) 500 g categories (African-Americans) Very low birthweight (whites) Very LBW (African-Americans) Low birthweight 1500-2500 g (whites) Low birthweight (African-Americans) Normal ≥2500 g (whites) Normal ≥2500 g (African-Americans)		 97% 94% 98% 97% 99.7% 99.7%	r = 0.99 r = 0.98 99.4% 98.5%	National Natality Survey, stratified random sample of live births with oversampling of LBW, 1980 (N=9,941)[17,6]
Apgar score (5 minutes)	Exact value			100%	North Carolina, stratified random sample of 42 hospitals, 1989 (N=395)[18]

	Exact value (high-risk) Exact value (low-risk)			94% 97%	Tennessee, stratified random sample of live births, 1989 (N=1,016 high-risk cases; N=700 low-risk controls)[2]
	Exact value			r = 0.96 (mean difference 0)	Single teaching hospital, random sample of live births, 1995 (N=99)[13]
	Exact value Grouped value (0-3, 4-6, 7-10) 0-3 (whites) 4-6 (whites) 7-10 (whites)		 94.4% 91.6% 99.6%	95% 99%	National Natality Survey, stratified random sample of live births with oversampling of LBW, 1980 (N=9,941)[17,6]
	Exact value <7 <9	 100% 99.6%	 99.7% 98.6%	97.5%, κ = 0.91 99.6%, κ = 0.84 98.3%, κ = 0.89	NE Ohio, random sample of deliveries at 20 hospitals, 1993-95 (N=33,616)[4]
Birth order	First, second, third, etc. (if not single birth)			>99%	Tennessee, stratified random sample of live births, 1989 (N=1,016 high-risk cases; N=700 low-risk controls)[2]
Gender of child	Male vs. female			>98%	Tennessee, stratified random sample of live births, 1989 (N=1,016 high-risk cases; N=700 low-risk controls)[2]
	Male (vs. female)	99.5%	99.5%	99.5%, κ = 0.99	NE Ohio, random sample of deliveries at 20 hospitals, 1993-95 (N=33,616)[4]
Conditions of the newborn	Anemia Birth injury Hyaline membrane Assisted vent <30 mins Assisted vent >30 mins Seizures	9% 1-2% 33% 10-15% 37% 5%	50% 83-100% 90% 12-30% 95% 38%		Tennessee, stratified random sample of live births, 1989 (N=1,016 high-risk cases; N=700 low-risk controls)[2]
	Seizures	36-37%	65-76%		Kentucky, all neonates with seizures reported by 5 Fayette County hospitals, 1985-89 (N=58)[26]
	Hyaline membrane	34%	97%		Missouri, all VLBW infants born to St.Louis residents, 1989,91,92 (N=976)[27]
Congenital anomalies	Overall (anomalies detected by 1 year)	10%			Georgia, all liveborn infants in metro Atlanta, 1995 (N=40,266)[28]

	Overall (anomalies detected by 2 years)	20%	61%		Virginia, all liveborn infants at 16-22 pilot hospitals, 1986 (N=10,034)[29]
	Overall		88%		New York, cases reported as having malformations on birth certificates, 1983-86 (N=11,418)[30]
	Fetal alcohol syndrome	9%	12%		Alaska, all potential cases captured by FAS surveillance program using 16 data sources, 1989-90 (N=630)[31]
	Fetal alcohol syndrome	11%	18%		Colorado, all potential cases captured by FAS surveillance program using multiple sources, 1992-94 (N=228)[32]
	Congenital heart defects	9%[c]			Wisconsin, all CHD cases living in Milwaukee and treated at one teaching hospital, 1997-99 (N=373)[33]
	Down syndrome (overall) Down syndrome (whites) Down syndrome (African-Americans)	34% 37% 17%	92% 93% not reported		Ohio, all infants reported as having Down syndrome on birth certificates (N =824) or by cytogenetic laboratories, 1970-81 (N=1,010)[34]
	Hydrocephalus Heart malformation Malformed genitalia Renal agenesis Cleft lip and/or palate Polydactyly/syndactyly Clubfoot	28% 5% 27% 0% 53% 10% 15%	50% 64% 50% 0% 75% 100% 43%		Tennessee, stratified random sample of live births, 1989 (N=1,016 high-risk cases with birthweight <1500 g or neonatal death)[2]
	Spina bifida Rectal atresia or stenosis Esophageal atresia Omphalocele/gastroschisis Cleft lip and/or palate Clubfoot Diaphragmatic hernia Down syndrome Overall (anomalies detected by 1 year) Overall (anomalies detectable at birth)	40% 10% 12% 47% 38% 22% 33% 19% 14% 28%	100% 75% 67% 100% 98% 82% 64% 37% 77%		Georgia, all liveborn infants in metro Atlanta, 1989-90 (N=76,862)[35]

	Cleft lip and/or palate	66%			Colorado, all infants with clefts reported to Registry for Children with Special Needs, 1989 (N=99)[36]
	Anencephaly Spina bifida Microcephalus Hydrocephalus Transposition of great vessels Tetralogy of Fallot Ventricular septal defect Atrial septal defect Endocardial cushion defect Valve stenosis and insufficiency Hypoplastic left heart Coarctation of aorta Choanal atresia Lung agenesis, hypoplasia, dysplasia Colorectal atresia or stenosis Tracheoesophageal atresia Hypospadias/epispadias Cystic kidney disease Omphalocele/gastroschisis Cleft lip and/or palate Clubfoot Missing extremity Diaphragmatic hernia Down syndrome	64% 25% 2% 13% 10% 7% 1% 0% 4% 4% 11% 3% 0% 7% 6% 30% 6% 10% 0% 10-47% 9% 15% 6% 26%	69% 75% 20% 56% 44% 100% 40% 0% 50% 100% 100% 100% 29% 67% 100% 92% 67% 71-87% 79% 88% 100% 94%		California, all liveborn infants in SF Bay area, 1983 (N=66,481)[37]

	Craniofacial[d] Anencephaly Limbs[d] Spine and trunk skeleton[d] Pulmonary[d] Cardiovascular[d] Gastrointestinal[d] Skin, hair, muscle[d] Urogenital[d] Chromosomal[d]	6% 0% 5% 3% 7% 7% 4% 3% 6% 8%			Dallas, all deliveries at one public hospital, 1977-80 (N>28,000; N=423 with congenital anomalies)[38]

[a] All statistics shown are based on a denominator of at least 14 cases.
[b]This quantity is estimated from the data reported in the manuscript.
[c]Sensitivities ranged from 0% to 19% across nine hospitals, with a significant association between hospital size and identification of congenital heart disease (r=0.67).
[d] Sensitivities are also reported for subtypes of anomalies, although typically not for specific anomalies.

TABLE REFERENCES

1. Baumeister L, Marchi K, Pearl M, Williams R, Braveman P. The validity of information on "race" and "Hispanic ethnicity" in California birth certificate data. *Health Services Research* 2000; 35:869-883.
2. Piper JM, Mitchel EF, Jr., Snowden M, Hall C, Adams M, Taylor P. Validation of 1989 Tennessee birth certificates using maternal and newborn hospital records. *American Journal of Epidemiology* 1993; 137:758-768.
3. Reichman NE, Hade EM. Validation of birth certificate data: A study of women in New Jersey's HealthStart program. *Annals of Epidemiology* 2001; 11:186-193.
4. DiGiuseppe DL, Aron DC, Ranbom L, Harper DL, Rosenthal GE. Reliability of birth certificate data: a multi-hospital comparison to medical records information. *Maternal and* Child Health Journal 2002; 6:169-179.
5. Schoendorf KC, Parker JD, Batkhan LZ, Kiely JL. Comparability of the birth certificate and 1988 Maternal and Infant Health Survey. *Vital & Health Statistics - Series 2: Data Evaluation & Methods Research* 1993; 116:1-19.
6. Fingerhut LA, Kleinman JC. Comparability of reporting between the birth certificate and the 1980 National Natality Survey. *Vital & Health Statistics - Series 2: Data Evaluation & Methods Research* 1985; 93:1-34.
7. Braveman P, Cubbin C, Marchi K, Egerter S, Chavez G. Measuring socioeconomic status/position in studies of racial/ethnic disparities: Maternal and infant health. *Public Health Reports* 2001; 116:449-463.
8. Marshall EG, Gensburg LJ, Roth GB, Davidson GK, Dlugosz LJ. Comparison of mother's occupation and industry from the birth certificate and a self-administered questionnaire. *Journal of Occupational Medicine* 1992; 34:1090-1096.
9. Shaw GM, Malcoe LH, Croen LA, Smith DF. An assessment of error in parental occupation from the birth certificate. *American Journal of Epidemiology* 1990; 13:1072-1079.

10. Braveman P, Pearl M, Egerter S, Marchi K, Williams R. Validity of insurance information on California birth certificates. *American Journal of Public Health* 1998; 88:813-816.
11. Dobie SA, Baldwin LM, Rosenblatt RA, Fordyce MA, Andrilla CHA, Hart LG. How well do birth certificates describe the pregnancies they report? The Washington State experience with low-risk pregnancies. *Maternal and Child Health Journal* 1998; 2:145-154.
12. Adams M. Validity of birth certificate data for the outcome of the previous pregnancy, Georgia, 1980-1995. *American Journal of Epidemiology* 2001; 154:883-888.
13. Costakos DT, Love LA, Kirby RS. The computerized perinatal database: are the data reliable? *American Journal of Perinatology* 1998; 15:453-459.
14. Emery ES III, Eaton A, Grether JK, Nelson KB. Assessment of gestational age using birth certificate data compared with medical record data. *Paediatric and Perinatal Epidemiology* 1997; 11:313-321.
15. McDermott J, Drews C, Green D, Berg C. Evaluation of prenatal care information on birth certificates. *Paediatric and Perinatal Epidemiology* 1997; 11:105-121.
16. Roohan PJ, Josberger RE, Acar J, Dabir P, Feder HM, Gagliano PJ. Validation of birth certificate data in New York State. *Journal of Community Health* 2003; 28:335-346.
17. Penrod JR, Lantz PM. Measurement error in prenatal care utilization: Evidence of attenuation bias in the estimation of impact on birth weight. *Maternal and Child Health Journal* 2000; 4:39-52.
18. Buescher PA, Taylor KP, Davis MH, Bowling JM. The quality of the new birth certificate data: A validation study in North Carolina. *American Journal of Public Health* 1993; 83:1163-1165.
19. Denk CE. Validation of EBC medical risk factors using longitudinal linkage, New Jersey 1996-2001. Presented at the 130th Annual Meeting of the American Public Health Association; November 11, 2002.
20. Clark K, Fu CM, Burnett C. Accuracy of birth certificate data regarding the amount, timing, and adequacy of prenatal care using prenatal clinic medical records as referents. *American Journal of Epidemiology* 1997; 145:68-71.
21. Forrest JD, Singh S. Timing of prenatal care in the United States: How accurate are our measurements? *Health Services Research* 1987; 22:235-253.
22. Dietz PM, Adams MM, Kendrick JS, Mathis MP. Completeness of ascertainment of prenatal smoking using birth certificates and confidential questionnaires: variations by maternal attributes and infant birth weight. PRAMS Working Group. Pregnancy Risk Assessment Monitoring System. *American Journal of Epidemiology* 1998; 148:1048-1054.
23. Parrish KM, Holt VL, Connell FA, Williams B, LoGerfo JP. Variations in the accuracy of obstetric procedures and diagnoses on birth records in Washington State, 1989. *American Journal of Epidemiology* 1993; 138:119-127.
24. Walker DS, Schmunk SB, Summers L. Do birth certificate data accurately reflect the number of CNM-attended births? An exploratory study. *Journal of Midwifery & Women's Health* 2004; 49:443-448.
25. Green DC, Moore JM, Adams MM, Berg CJ, Wilcox LS, McCarthy BJ. Are we underestimating rates of vaginal birth after previous cesarean birth? The validity of delivery methods from birth certificates. *American Journal of Epidemiology* 1998; 17:581-586.
26. Lanska MJ, Lanska DJ, Baumann RJ. A population-based study of neonatal seizures in Fayette County, Kentucky: comparison of ascertainment using different health data systems. *Neuroepidemiology* 1995; 14:278-285.
27. Hamvas A, Kwong P, DeBaun M, Schramm W, Cole FS. Hyaline membrane disease is underreported in a linked birth infant-death certificate database. *American Journal of Public Health* 1998; 88:1387-1389.
28. Honein MA, Paulozzi LJ. Birth defects surveillance: assessing the "gold standard". *American* Journal of Public Health 1999; 89:1238-1240.

29. Marazita ML, Bodurtha JN, Corey L, Rogers A, Barbosa CE, Funkhouser L *et al.* Development of the Virginia Congenital Anomalies Reporting and Education System (VaCARES): two pilot projects. *Southern Medical Journal* 1992; 85:608-615.
30. Olsen CL, Polan AK, Cross PK. Case ascertainment for state-based birth defects registries: characteristics of unreported infants ascertained through birth certificates and their impact on registry statistics in New York state. *Paediatric and Perinatal Epidemiology* 1996; 10:161-174.
31. Egeland GM, Perham-Hester KA, Gessner BD, Ingle D, Berner JE, Middaugh JP. Fetal alcohol syndrome in Alaska, 1977 through 1992: an administrative prevalence derived from multiple data sources. *American Journal of Public Health* 1998; 88:781-786.
32. Miller LA, Shaikh T, Stanton C, Montgomery A, Rickard R, Keefer S *et al.* Surveillance for fetal alcohol syndrome in Colorado. *Public Health Reports* 1995; 110:690-697.
33. Cronk CE, Malloy ME, Pelech AN, Miller RE, Meyer SA, Cowell M *et al.* Completeness of state administrative databases for surveillance of congenital heart disease. *Birth Defects Research* 2003; 67:597-603.
34. Johnson KM, Huether CA, Hook EB, Crowe CA, Reeder BA, Sommer A *et al.* False-positive reporting of Down syndrome on Ohio and New York birth certificates. *Genetic Epidemiology* 1985; 2:123-131.
35. Watkins ML, Edmonds L, McClearn A, Mullins L, Mulinare J, Khoury M. The surveillance of birth defects: The usefulness of the revised US Standard Birth Certificate. *American Journal* of Public Health 1996; 86:731-734.
36. Amidei RL, Hamman RF, Kassebaum DK, Marshall JA. Birth prevalence of cleft lip and palate in Colorado by sex distribution, seasonality, race/ethnicity, and geographic variation. *Special Care in Dentistry* 1994; 14:233-240.
37. Hexter AC, Harris JA, Roeper P, Croen LA, Krueger P, Gant D. Evaluation of the hospital discharge diagnoses index and the birth certificate as sources of information on birth defects. *Public Health Reports* 1990; 105:290-307.
38. Snell LM, Little BB, Knoll KA, Johnston WL, Jr., Rosenfeld CR, Gant NF. Reliability of birth certificate reporting of congenital anomalies. *American Journal of Perinatology* 1992; 9:219-222.

Section 5: Are Postpartum Readmissions a Valid Measure of Obstetric Quality of Care?

SUMMARY

Objective: To assess the validity of risk-adjusted postpartum maternal readmission rates based on California's Patient Discharge Data Set as a measure of obstetric quality of care.

Research Design: Retrospective cohort.

Subjects: We randomly sampled 1,611 deliveries from 52 of the 267 nonfederal hospitals that performed more than 678 eligible deliveries in California in 1992-1993, oversampling hospitals with significantly more or fewer readmissions than expected.

Measures: Using data abstracted from medical records by experienced obstetric nurses, we compared in-hospital postpartum complication rates, processes of care, and severity of illness at readmission across hospitals with low, intermediate, and high risk-adjusted readmission rates.

Results: Most post-cesarean complications were more prevalent at high-readmission than at low-readmission hospitals (laceration, 9.4% versus 0.7%; wound infection, 5.9% versus 0.0%; endometritis, 13.3% versus 2.5%; mean hematocrit drop 4.8% versus 3.9%). There were no significant differences in labor management, but women at high-readmission hospitals were less likely to have operative vaginal delivery (1.5% versus 16%) or scheduled repeat cesarean (17% versus 58%), and more likely to have vaginal birth after cesarean (60% versus 28%), than women at low-readmission hospitals. Mean postpartum length-of-stay was shorter at low-readmission hospitals than elsewhere (vaginal, 28 versus 31-32 hours; cesarean, 70 versus 75-82 hours). The percentage of women in the highest acuity level at readmission was similar at low (62%), intermediate (64%), and high-readmission (52%) hospitals.

Conclusions: Although we could not establish whether low-readmission hospitals provide better care than high-readmission hospitals, low risk-adjusted postpartum readmission rates are associated with a more aggressive management style and fewer post-cesarean complications.

INTRODUCTION

Readmissions are a widely used measure of quality of care for medical and surgical conditions,[1,2] in part because they can be ascertained using linked data from hospital discharge or claims databases. A conceptual model linking readmissions to antecedent quality of care and premature discharge has been proposed.[3] Patients hospitalized for chronic conditions such as heart failure and obstructive lung disease have an especially high risk of readmission;[4] several randomized controlled trials have identified inpatient interventions that reduce this risk.[5,6,7,8] In observational settings, some but not all studies have shown that readmitted patients experienced poorer processes of care during their prior hospitalization than patients who were not readmitted, after adjusting for differences in severity of illness.[9,10]

This study was conducted to validate the use of risk-adjusted postpartum maternal readmission rates as an indicator of the quality of inpatient obstetric care in California. Because content validity had already been addressed through the deliberations of an expert clinical advisory panel, we focused on testing criterion and construct validity. These validation analyses required independent recoding of medical records by coding professionals, supplemented by detailed review of medical and nursing management by experienced obstetric nurses.

Criterion validity describes the extent to which a measure being evaluated agrees with a "gold standard" or better measure of the same phenomenon. Accordingly, we hypothesized that if risk-adjusted postpartum readmission rates are a valid quality indicator, then these rates should be relatively robust to differences in risk-adjustment methodology. Hospitals with high rates when risk-adjustment is performed using administrative data should also have high rates when risk-adjustment is performed using more detailed clinical data.

We then performed two sets of analyses to address construct validity. Our first construct was that readmissions often result from in-hospital complications that are not recognized or not appropriately treated at discharge. If this construct is correct, then hospitals with high risk-adjusted rates of postpartum readmission should also have high rates of important in-hospital postpartum complications. Our second construct was that poor clinical outcomes, such as complications requiring readmissions, often result from poor processes of care. If this construct is correct, then hospitals with high risk-adjusted rates of postpartum readmission should manage labor, delivery, and the puerperium in ways that predispose to complications.

Finally, we evaluated the possibility of confounding due to variation in the clinical indications or thresholds for postpartum readmission across hospitals. In the absence of confounding, patients who are readmitted after delivery at hospitals with high risk-adjusted readmission rates should be just as sick as patients who are readmitted after delivery elsewhere. Differences in severity of illness at readmission would suggest that hospitals with high readmission rates may be less selective in whom they readmit, or may see more vulnerable patients who would have difficulty managing their complications in the outpatient setting.

METHODS

The general methods for this study are described in Section 2. In this section, we describe only

those aspects of the methods that were specifically related to evaluating the validity of postpartum maternal readmission rates as an indicator of hospital quality.

In defining readmissions, we excluded "direct" transfers to *other* acute care facilities. Transfers were labeled as "direct" if the discharge disposition was "acute hospital" with an discharge-to-readmit interval of 0-1 days, or if the disposition was "other facility" or "routine" with a discharge-to-readmit interval of 0 days. We also excluded readmissions with principal diagnoses unlikely to be related to inpatient obstetric care: malignancy, mental disorders (290.xx-319, 648.3x-648.4x, V71.0x), cardiomyopathy (425.x), cardiac dysrhythmias and failure (427.xx-428.x), cerebrovascular disease (430-437.x, 674.0x), appendicitis (540.x-543.x), enteritis and colitis (001.x-009.x, 555.x-558.x), cholelithiasis and cholecystitis (574.xx-575.xx), diseases of pancreas (577.x), calculus of kidney and ureter (592.x), disorders of breast (610.x-611.xx, 675.xx-676.xx), injuries (800.xx-959.xx), sterilization (V25.2), and rehabilitation (V57.xx). This list was not intended to be exhaustive; other obviously unrelated diagnoses would have been excluded if they had been represented. We captured all eligible readmissions, regardless of location, using the patient's encrypted social security number (confirmed by at least partial match on date of birth) to link across hospitalizations.

After selecting a stratified random sample of vaginal and cesarean deliveries from acute care hospitals in California, as described in Section 2, we asked each participating hospital to photocopy each sampled record and associated prenatal records, if available. Each record was reviewed by one of four experienced Accredited Record Technicians or Certified Coding Specialists, who abstracted demographic and prenatal data, blinded to both the original discharge abstract and the hospital's readmission rate. A regional coding authority tested these individuals before they were hired, trained and supervised them, and verified at least 10% of each abstractor's records to ensure at least 95% accuracy. Discrepancies were resolved through collective review of appropriate references. One of two experienced obstetric nurses (also blinded to hospital-reported data) then abstracted more complex clinical data.[11] To enhance data accuracy, we examined the univariate distribution of each variable and corrected illogical values by re-reviewing medical records.

We evaluated the criterion validity of risk-adjusted postpartum readmission rates based on OSHPD's Patient Discharge Data Set by estimating risk-adjustment models using the recoded ICD-9-CM diagnoses and reabstracted clinical data described above. Separate models were estimated for vaginal and cesarean deliveries, to permit direct comparison of criterion validity between women with different modes of delivery. Because our validation sample was much smaller than the original sample used to select risk factors, we retained the risk factors identified through the modeling procedure described in Section 2. We then re-estimated these models on our validation sample to obtain sample-specific coefficient estimates. To avoid convergence problems and reversals of the direction of association, up to three risk factors were removed from each model. Two dummy variables were added to each model to estimate the adjusted odds ratio for postpartum readmission at hospitals that were flagged as having significantly more or fewer readmissions than expected (relative to non-outlier hospitals). For each risk factor, we substituted the version based on recoded ICD-9-CM diagnoses or reabstracted demographic data for the version based on hospital-reported data. The only exceptions to this substitution were insurance status, which was not reabstracted, and the number of antepartum hospitalizations,

which could only be determined from the Patient Discharge Data Set.

We then added or substituted several clinical risk factors that could not be captured fully using ICD-9-CM codes alone: previous cesarean, multiple gestation, premature delivery (<28, 28-32, or 32-37 weeks of gestation), preeclampsia or eclampsia (clinically defined), thick meconium at delivery, crash cesarean (cesarean deliveries only), insulin-requiring diabetes mellitus, antepartum fever, maternal obesity (body mass index >95th percentile [42.1]) or overweight (body mass index between the 85th and 95th percentiles [33.7-42.1]) at admission, grand multiparity (e.g., at least 4 prior vaginal deliveries), fetal macrosomia (birth weight >4000 grams), psychiatric disease, third trimester hemorrhage (vaginal deliveries only), and antepartum anemia (hematocrit <33%, vaginal deliveries only). Two dummy variables were added to each model to estimate the adjusted odds ratio for postpartum readmission at hospitals that were flagged as having significantly more or fewer readmissions than expected (relative to non-outlier hospitals). In a confirmatory analysis, we omitted these two variables and used the resulting models to re-estimate each patient's risk of readmission based on recoded and clinical data. These probabilities were summed at the hospital level to estimate the expected number of readmissions for each hospital. The indirectly standardized readmission ratio for each hospital sampling stratum was estimated as the ratio of observed to expected readmissions. The patient-level c statistic, and the hospital-level weighted Pearson correlation coefficient between observed and expected readmission rates, were estimated as measures of overall model performance.

We evaluated the construct validity of risk-adjusted postpartum readmission rates as a measure of inpatient quality of care, based on Donabedian's construct of structure, process, and outcomes.[12] We hypothesized that postpartum outcome measures would be associated with each other, and with process measures of quality at the hospital level. Accordingly, we compared the prevalence of inpatient postpartum complications and relevant processes of care across hospitals with fewer readmissions than expected ($p<0.01$), more readmissions than expected ($p<0.01$), and neither. In a sensitivity analysis, we used $p<0.10$ as an alternative threshold for classifying hospitals. These comparisons across hospital sampling strata were performed using the svytab procedure in STATA Release 6, which provides robust variance estimates taking into account the oversampling of cesarean deliveries and the clustering of observations within hospitals.

Risk-adjusted readmission rates may be confounded by unmeasured variation in either severity of illness or physicians' admitting practices across hospitals. Accordingly, we also defined definite and acceptable criteria for postpartum readmission among women with the four most common indications: hemorrhage and retained products, endometritis, wound infection, and urinary tract infection. These criteria were based on key clinical findings that we abstracted from clinic and emergency room notes at presentation, including symptoms (e.g., fever, abdominal/ pelvic pain, vaginal discharge, urinary symptoms, wound discharge or separation, vomiting), prior outpatient therapies (e.g., antibiotics, wound care), physical findings (e.g., temperature, heart rate, blood pressure, abdominal/pelvic examination, wound examination), and laboratory values (e.g., hematocrit, leukocyte count, pelvic ultrasound). We then compared the percentage of readmitted women with definite (or any acceptable) criteria for readmission across hospitals with fewer readmissions than expected ($p<0.01$), more readmissions than expected ($p<0.01$), and neither. In a sensitivity analysis, we used $p<0.10$ as an alternative threshold for classifying hospitals. These comparisons across hospital sampling strata were again performed using the

svytab procedure in STATA Release 6.

Because of the complex sample structure, all analyses were weighted (unless otherwise noted). The weight was defined as the inverse of the sampling probability, which was calculated by multiplying the probability of sampling a specific hospital by the probability of sampling an individual within that hospital. The weights were adjusted to reflect both nonsubmitted records and records that were later classified as ineligible.

RESULTS

We received 1,614 of the 1,662 delivery records (97.1%) requested from the 52 participating hospitals, and 469 of the 493 requested readmission records (95.1%). Three of the former and four of the latter records did not qualify for the study; hence 1,611 delivery (471 primary cesarean, 319 repeat cesarean, 821 vaginal) and 465 readmission records were abstracted.

Table 1 compares the overall performance of risk-adjustment models based solely on hospital discharge data and similar models based on recoded and reabstracted data with additional clinical detail. Specific parameter estimates from these models are shown in Appendix Tables 1 through 4. Overall model discrimination, as measured by the c statistic (e.g., the area under a receiver operating characteristic curve), improved moderately when recoded and clinical data were substituted for hospital-reported data. Similarly, the proportion of variation in hospital readmission rates explained by the risk-adjustment model increased, but remained under 4%. The rank correlation between hospital-level predicted readmission rates from these models was moderate (r=0.38). The adjusted odds ratios for readmission after vaginal delivery were very similar between these models, both for hospitals with significantly fewer complications than expected (OR=0.42, 95% CI 0.35-0.51 based on hospital-reported data versus OR=0.43, 95% CI 0.35-0.52 based on recoded and clinical data) and for hospitals with significantly more complications than expected (OR=1.21, 95% CI 1.09-1.33 based on hospital-reported data versus OR=1.33, 95% CI 1.19-1.47 based on recoded and clinical data). Similarly, the ratios of observed to expected readmissions at outlier hospitals were essentially unchanged by better risk-adjustment using clinical data. The overall difference in the risk-adjusted readmission rate between low and high outlier hospitals decreased slightly, from 0.586% to 0.563%, with better risk-adjustment using clinical data.

Table 2 shows the prevalence of various in-hospital, postpartum complications at hospitals with fewer readmissions than expected, more readmissions than expected, and neither fewer nor more readmissions than expected (hereafter referred to as low, intermediate, and high-readmission hospitals). Only complications documented in at least 12 women are shown. Weighted prevalence estimates can be extrapolated to the entire population of women who were delivered at acute care, non-federal hospitals with active obstetric services in California.

For most in-hospital post-cesarean complications, there was a statistically or marginally significant ($p<0.10$) trend toward higher prevalence at either high or intermediate-readmission hospitals, relative to low-readmission hospitals. For example, cesarean-related lacerations were documented in 9.4% of eligible patients at high-readmission hospitals versus 0.7% at low-readmission hospitals. Cesarean wound infections were documented in 5.9% of eligible patients

at high-readmission hospitals versus 0.0% at low-readmission hospitals. Endometritis was documented in 13.3% of cesarean patients at high-readmission hospitals versus 2.5% at low-readmission hospitals. A similar difference was noted for post-cesarean fever (33.5% versus 20.4%), an objective finding that is less susceptible to documentation bias than endometritis. Although the prevalence of postpartum hemorrhage did not differ significantly across hospital strata, the estimated blood loss during and the measured hematocrit drop after cesarean delivery were greater at high-readmission than at low-readmission hospitals (832 versus 648 ml, and 4.8% versus 3.9%, respectively). Prevalence differences across hospital strata were less consistent for complications after vaginal delivery.

Table 3 shows no significant differences in labor management across hospitals stratified according to their readmissions experience. However, there were striking differences in delivery management, in that women at high-readmission hospitals were less likely to have operative vaginal deliveries than women at other hospitals (1.5% versus 16.1% at low-readmission and 13.8% at intermediate-readmission hospitals). Among these women, forceps were used in most cases at high-readmission hospitals (74%), whereas vacuum assistance was used in most cases at other hospitals (69%-70%). Mediolateral episiotomies were almost never performed at high-readmission hospitals, but were performed occasionally (2.9%-5.3%) at other hospitals. At low-readmission hospitals, 16% of eligible patients were not counseled about a trial of labor, and 47% reportedly refused despite counseling, resulting in an unusually high rate of scheduled repeat cesarean deliveries (58% versus 17% at low-readmission and 35% at intermediate-readmission hospitals) and a low rate of vaginal birth after cesarean (28% versus 60% at low-readmission and 51% at intermediate-readmission hospitals).

With respect to postpartum care, low-readmission hospitals were more likely than other hospitals to have a documented nurse inspection of the cesarean wound at discharge, but less likely to have a documented physician inspection (94% and 59% at low-readmission hospitals, respectively, versus 85% and 76% at high-readmission and 85% and 71% at intermediate-readmission hospitals). Mean postpartum length of stay was also shorter at low-readmission than at intermediate-readmission hospitals for both vaginal and cesarean deliveries, consistent with the finding in Table 2 that in-hospital postpartum complication rates were less frequent at low-readmission hospitals than at other hospitals.

Finally, Table 4 shows the percentage of readmissions that met definite or any criteria for readmission, based on nurses' abstraction of key clinical findings that physicians consider in determining whether to readmit a postpartum patient. Overall, the percentage of women in the highest acuity level at readmission was similar at low-readmission (62%), intermediate-readmission (64%), and high-readmission (52%) hospitals. Only for postpartum endometritis was there evidence that women readmitted at high-readmission hospitals were less severely ill than women readmitted at low or intermediate-readmission hospitals (i.e., 42% met definite criteria for readmission, versus 56% and 62%, respectively). This difference disappeared when looser criteria for readmission were applied or when postpartum endometritis was aggregated with other complications that often necessitate readmission. Sensitivity analyses using $p<0.10$ as an alternative threshold for classifying hospital readmission rates generated results similar to those presented in Tables 2-4, except that the difference between low and high-readmission hospitals in the overall percentage of women with any criteria for readmission (83% versus 73%,

respectively) achieved statistical significance (Table 5).

DISCUSSION

The current study represents the first comprehensive effort to evaluate the validity of risk-adjusted postpartum maternal readmission rates as an indicator of the quality of inpatient obstetric care. In another component of the same study,[13] we showed that inter-hospital variation in coding does not significantly bias risk-adjustment of readmission rates, because high-readmission hospitals are not systematically less likely to report risk factors than low-readmission hospitals. In this Section, we performed four sets of analyses to address the validity of risk-adjusted postpartum maternal readmission rates as an indicator of the quality of inpatient obstetric care. First, we established that it is possible to construct better risk-adjustment models for postpartum readmissions using additional clinical data elements that can only be obtained through medical records review by trained professionals. However, these improved models do not explain the superior performance of hospitals with significantly fewer than expected readmissions, nor the inferior performance of hospitals with significantly more than expected readmissions. By several metrics, the performance gap between outlier hospitals on either end of the distribution remains roughly constant despite better risk-adjustment, suggesting that measures of risk-adjusted readmissions based only on OSHPD data have criterion validity.

Second, we confirmed that women who were delivered at high-readmission hospitals were also more likely to experience several types of in-hospital postpartum complications than women who were delivered at low-readmission hospitals. These complications included cesarean-related laceration, wound infection, endometritis, and postpartum fever. In addition, women who underwent cesarean delivery at high-readmission hospitals lost more blood, and experienced a greater fall in hematocrit, than women who underwent cesarean delivery at low-readmission hospitals. These findings support the validity of risk-adjusted postpartum maternal readmission rates as a quality indicator, at least for cesarean delivery.

Our third set of analyses demonstrated substantial differences in processes of care across low, intermediate, and high-readmission hospitals. However, these differences are difficult to interpret because they do not follow a consistent pattern. With input from a clinical advisory panel, we identified several physician and nurse behaviors that may increase the risk of postpartum infection, including artificially rupturing membranes, failing to intervene after 12 hours of rupture, performing multiple vaginal examinations after rupture, internal fetal monitoring, failing to give timely antibiotics to febrile women or to women undergoing cesarean delivery, inserting Foley catheters before vaginal delivery, and failing to remove them promptly. None of these behaviors was more prevalent at high-readmission than at low-readmission hospitals. However, physicians at high-readmission hospitals used vacuum assistance less often than physicians at low-readmission hospitals; they were also more likely to offer a trial of labor to eligible women after a prior cesarean, less likely to document refusal of that offer, and more likely to have successful vaginal births after cesarean. Postpartum hospital stays were longer, on average, at intermediate-readmission (and nonsigificantly at high-readmission) than at low-readmission hospitals, consistent with the observed difference in in-hospital complication rates. We conclude that low risk-adjusted postpartum readmission rates are associated at the hospital level with a more interventional style of delivery management, although we cannot determine

whether this relationship is causal.

Finally, we confirmed that variation in readmission practices is unlikely to explain the observed difference in risk-adjusted postpartum maternal readmission rates between low and high-readmission hospitals. Only for postpartum endometritis did we find evidence that high-readmission hospitals admit less severely ill women than low-readmission hospitals. Although the percentage of readmitted women who met definite clinical criteria for readmission varied from 52% to 64% across hospital strata, this variation could not account for the three-fold difference in risk-adjusted readmission rates across strata.

Although we are not aware of any previous studies of the validity of risk-adjusted postpartum readmission rates as an indicator of inpatient quality of care, our approach is consistent with how other researchers have analyzed this outcome. For example, several authors and meta-analysts have used postpartum maternal readmissions within 6 weeks as an outcome in evaluating the effects of shorter hospital stays after delivery[14,15,16,17] and related changes in the organization and financing of perinatal care.[18,19] Pregnancy-related complications and prenatal emergency room visits and admissions were significant predictors of postpartum readmissions in an analysis of Medicaid claims from three California counties, but the authors found no effect of Medicaid managed care.[20] The strongest risk factors for maternal readmission in Canada were in-hospital postpartum complications such as puerperal infection and postpartum hemorrhage, plus selected antepartum complications such as polyhydramnios, hypertension, and diabetes.[21] Antepartum transfusions may flag women who are at especially high risk.[22]

In interpreting our findings, several limitations should be considered. First, our review of medical records may have failed to capture some in-hospital complications or some clinical findings that justified readmission. However, we selected nurses with experience in obstetrics, trained them thoroughly using specific written guidelines, monitored them carefully, and gave them as much time as needed to abstract each record. Nonetheless, our abstractors were only able to abstract information that was documented in the medical record; physician and nursing documentation was manifestly inadequate in some cases. Second, because this study was designed to validate a published report on risk-adjusted outcomes, our medical records came from 1992 and 1993. Readmission practices may have changed during the past decade, such that readmission rates may now be a better or poorer indicator of quality of care. Similarly, ICD-9-CM coding practices may have improved so much that coding errors and omitted clinical risk factors may have even less impact on model performance and hospital classification. Our exclusion of very low-volume hospitals from the sampling frame may limit the generalizability of our findings. Fortunately, these hospitals accounted for fewer than 10% of deliveries performed in California hospitals during the study period. Finally, despite the fact that our sample was designed to provide 80% power to find meaningful differences in key processes and outcomes of care between low, intermediate, and high-readmission hospitals, postpartum readmissions remain a rare outcome that may be subject to excessive random variation.[23]

Despite these limitations, our findings have important implications for assessing obstetric quality of care. It appears that high-readmission hospitals see more postpartum complications (especially post-cesarean complications), both before and after discharge, than low-readmission hospitals. However, the reasons for these differences remain unclear. We cannot say whether

low-readmission hospitals actually provide better care than high-readmission hospitals, although it is certainly plausible that they do. Given that low readmission rates appear to be a marker for a more aggressive style of delivery management, we suspect that there are major differences in bedside care that could not be ascertained by abstracting explicit process measures from medical records. Future studies should consider other process measures, and apply both implicit and explicit review to capture more subjective dimensions of quality.

Table 1. Measures of risk-adjustment performance and accuracy, using either hospital-reported data or professionally recoded and reabstracted data, from 52 California hospitals previously identified as having more postpartum readmissions than expected, fewer readmissions than expected, or neither

Measure	**Hospital-reported ICD-9-CM codes and demographic data**	**Recoded ICD-9-CM codes plus reabstracted clinical data**
Adjusted odds of readmission after vaginal delivery		
Hospitals with fewer (p<0.01) readmissions than expected	0.42 (0.35-0.51)	0.43 (0.35-0.52)
Neither	1.0 (reference)	1.0 (reference)
Hospitals with more readmissions than expected	1.21 (1.09-1.33)	1.33 (1.19-1.47)
Adjusted odds of readmission after cesarean delivery		
Hospitals with fewer (p<0.01) readmissions than expected	0.40 (0.30-0.52)	0.56 (0.43-0.73)
Neither	1.0 (reference)	1.0 (reference)
Hospitals with more readmissions than expected	1.94 (1.67-2.26)	1.91 (1.63-2.24)
Ratio of observed to expected number of readmissions		
Hospitals with fewer (p<0.01) readmissions than expected	0.452	0.507
Neither	1.008	0.996
Hospitals with more readmissions than expected	1.344	1.365
Overall difference in risk-adjusted readmission rate between low and high-outlier hospitals	0.586%	0.563%
Weighted hospital-level Pearson correlation between observed and expected readmissions		
Number of readmissions	0.956	0.957
Readmission rate	0.105	0.197
C statistic		
Vaginal deliveries	0.587	0.619
Cesarean deliveries	0.631	0.651

Table 2. In-hospital postpartum complication rates, based on nurses' review of medical records, at California hospitals with more postpartum readmissions than expected, fewer readmissions than expected, and neither

Postpartum complication	**Type of delivery V: vaginal C: cesarean**	**No. women with complication***	**Weighted Prevalence (%)**		
			Hospitals with fewer (p<0.01) readmissions (N=413)	**Other Hospitals (N=828)**	**Hospitals with more (p<0.01) readmissions (N=370)**
Perineal laceration (any degree)	V	301	24.4%†	38.6%†	30.2%
High vaginal/cervical laceration	V	84	4.5%‡	11.2%‡§	4.0%§
Other laceration	C	32	0.7%**	1.8%†	9.4%**†
Hemorrhage	All	75	3.5%	2.3%	1.4%
Vaginal delivery	V	47	4.5%	2.0%	1.4%
Cesarean delivery	C	28	0.6%††	3.4%††	1.3%
With Transfusion	All	17	0.6%	0.2%	0.2%
Dilatation & curettage	All	14	0.6%	0.4%	1.1%
Urinary tract infection	All	16	0.6%	0.3%	0.3%
Wound infection	All	41	1.3%	1.0%	1.7%
Cesarean delivery	C	39	0.0%†‡	2.8%†	5.9%‡
Discharged with wound abnormality per RN note					
Marginal or definite	C	40	0.2%†††‡‡	1.1%††	1.7%‡‡
Definite	C	22	0.0%††	0.4%††	0.5%

* Only complications that occurred in at least 12 women are shown.
† The percentages designated by this symbol are significantly different by pairwise comparison, p<0.01.
‡ The two percentages in the same row that are marked with this symbol are significantly different by pairwise comparison, p<0.05.
§ The two percentages in the same row that are marked with this symbol are significantly different by pairwise comparison, p<0.05.
** The two percentages in the same row that are marked with this symbol are significantly different by pairwise comparison, p<0.005.
†† The two percentages in the same row that are marked with this symbol are marginally different by pairwise comparison, p<0.10.
‡‡ The two percentages in the same row that are marked with this symbol are marginally different by pairwise comparison, p<0.10.

Discharged with wound abnormality per MD note					
Marginal or definite	C	49	0.8%	1.2%	2.3%
Definite	C	24	0.2%	0.5%	1.4%
Endometritis	All	59	1.1%	2.0%	3.4%
Cesarean delivery	C	49	2.5%††	5.6%	13.3%††
Fever (T_{max}>38.4 °C)	All	315	8.9%	9.4%	12.0%
Vaginal delivery	V	67	5.1%	5.8%	7.4%
Cesarean delivery	C	248	20.4%††	22.2%	33.5%††
High fever (T_{max}>38.9 °C)	All	28	0.8%	1.4%	1.5%
Cesarean delivery	C	22	1.3%	1.5%	2.7%
Any specified complication	All	777	32.9%§§	48.0%§§	42.2%
Vaginal delivery	V	437	35.3%§§	52.7%§§	41.6%
Cesarean delivery	C	340	25.6%**	31.2%**‡	44.8%‡
Estimated blood loss (ml)	C	784***	648**††	736††‡‡	832**‡‡
Hematocrit drop (%)	All	1116***	3.4	3.5	3.8
Vaginal delivery	V	449***	3.2	3.1	3.5
Cesarean delivery	C	667***	3.9††	4.4	4.8††

§§ The two percentages in the same row that are marked with this symbol are significantly different by pairwise comparison, p<0.001.
*** Number of cases with nonmissing data.

Table 3. In-hospital processes of care, based on nurses' review of medical records, at California hospitals with more postpartum readmissions than expected, fewer readmissions than expected, and neither

Process of care	Eligible population	No. women eligible for specified process of care	Weighted Prevalence (%)		
			Hospitals with fewer (p<0.01) readmissions (N=413)	Other Hospitals (N=828)	Hospitals with more (p<0.01) readmissions (N=370)
Labor management					
Spontaneous rupture of membranes Vaginal deliveries	All excluding elective C/S and ruptured membranes at admit Vaginal deliveries	815 815	26.0 49.0	26.2 46.7	24.1 45.3
Duration rupture of membranes (hrs)	All excluding elective cesareans	1210	7.7	8.2	9.8
Prolonged rupture of membranes (>12 hrs)	All excluding elective cesareans	1279	1.1[§]	3.9[§]	1.6
Pelvic exams after rupture of membranes	All excluding elective cesareans	1316	5.1	4.6	4.9
Internal monitoring	All excluding elective cesareans	1316	35.6	34.7	43.8
Induction of labor	All excluding elective cesareans	1316	13.5	9.9	13.2
Epidural analgesia	Vaginal deliveries	821	39.1	27.8	37.1
Spinal anesthesia	Cesarean deliveries	790	48.8[†]	38.5[**]	22.0[†**]
Delivery management					
Operative vaginal delivery Forceps alone Vacuum alone	All excluding elective cesareans without trial of labor	1316 141 141	16.1[*] 18[†] 70[†]	13.8[‡] 30[§] 69[**]	1.5[*‡] 74[†§] 23[†**]

[*] The two percentages in the same row that are marked with this symbol are significantly different by pairwise comparison, p<0.0001.
[†] The two percentages in the same row that are marked with this symbol are significantly different by pairwise comparison, p<0.05.
[‡] The two percentages in the same row that are marked with this symbol are significantly different by pairwise comparison, p<0.0005.
[§] The two percentages in the same row that are marked with this symbol are marginally different by pairwise comparison, p<0.10.

Episiotomy Mediolateral	Vaginal deliveries With episiotomy	821 424	64.6 5.3[†]	53.0 2.9[§]	50.3 0[†§]
Incision other than low transverse	Cesarean deliveries	790	0.7[†]	5.8[†§]	1.3[§]
Eligible for trial of labor Scheduled repeat C/S Vaginal birth	Prior cesarean, eligible for trial of labor	 319 319	 57.6[†**] 27.7[†**]	 34.8[†] 50.8[†]	 16.5[**] 60.4[**]
Reason scheduled repeat C/S Failure to counsel Refused trial of labor	Prior cesarean, eligible for trial of labor	 313 258	 15.7[†] 46.8[†**]	 11.9[**] 24.8[†]	 0.1[†**] 16.4[**]
Operative time (minutes)	Cesarean deliveries	779	37.3	43.3	43.5
Foley catheter before delivery	Vaginal deliveries	821	31.0[†]	13.7[†**]	29.4[**]
Antibiotics before or within 60 minutes after delivery	Febrile before delivery	102	58.4	92.5	85.6
Antibiotics before or within 60 minutes after delivery	Cesarean deliveries	790	84.8	81.7	82.0
Postpartum management					
Foley catheter time (minutes) (hours)	Vaginal deliveries Cesarean deliveries	200 615	101 23.0	81 22.6	204 20.6
Time from delivery to ambulation (hrs)	Cesarean deliveries	760	23.6	23.6	22.9
Nurse inspected wound at discharge	Cesarean deliveries	788	93.7[†**]	85.3[†]	85.1[**]
Physician inspected wound at discharge	Cesarean deliveries	789	58.8[†§]	70.8[§]	75.7[†]
Postpartum hospital stay (hrs)	Vaginal deliveries Cesarean deliveries	802 765	28.2[§] 69.8[††]	31.9[§] 74.7[††]	31.3 81.8

[**] The two percentages in the same row that are marked with this symbol are significantly different by pairwise comparison, p<0.05.
[††] The two percentages in the same row that are marked with this symbol are significantly different by pairwise comparison, p<0.005.

Table 4. Severity of illness at postpartum readmission among patients readmitted to California hospitals with more postpartum readmissions than expected (p<0.01), fewer readmissions than expected (p<0.01), and neither

Principal diagnosis at readmission	No. readmissions for specified principal diagnosis	Percentage of women who meet specified criteria		
		Hospitals with fewer (p<0.01) readmissions	Other Hospitals	Hospitals with more (p<0.01) readmissions
Definite criteria for readmission (high acuity)				
Postpartum hemorrhage and retained products	76	67%	62%	62%
Postpartum endometritis	191	56%*	62%†	42%*†
Postpartum wound infection (cesarean or episiotomy)	75	56%	40%	44%
Postpartum urinary tract infection	63	30%‡	60%‡	51%
Any of the above	337	62%	64%	52%
Any criteria for readmission (high or intermediate acuity)				
Postpartum hemorrhage and retained products	76	96%	95%	100%
Postpartum endometritis	191	65%	78%	65%
Postpartum wound infection (cesarean or episiotomy)	75	66%	43%	44%
Postpartum urinary tract infection	63	36%*	66%*	51%
Any of the above	337	75%	80%	71%

* The two percentages in the same row that are marked with this symbol are significantly different by pairwise comparison, p<0.05.
† The two percentages in the same row that are marked with this symbol are significantly different by pairwise comparison, p<0.05.
‡ The two percentages in the same row that are marked with this symbol are marginally different by pairwise comparison, p<0.10.

Table 5. Severity of illness at postpartum readmission among patients readmitted to California hospitals with more postpartum readmissions than expected (p<0.10), fewer readmissions than expected (p<0.10), and neither

Principal diagnosis at readmission	No. readmissions for specified principal diagnosis	Percentage of women who meet specified criteria		
		Hospitals with fewer (p<0.10) readmissions	Other Hospitals	Hospitals with more (p<0.10) readmissions
Definite criteria for readmission (high acuity)				
Postpartum hemorrhage and retained products	76	50%	68%	57%
Postpartum endometritis	191	50%[*]	76%[†*]	37%[†]
Postpartum wound infection (cesarean or episiotomy)	75	55%	31%	51%
Postpartum urinary tract infection	63	48%	63%	49%
Any of the above	337	60%	69%[*]	51%[*]
Any criteria for readmission (high or intermediate acuity)				
Postpartum hemorrhage and retained products	76	99%	93%	100%
Postpartum endometritis	191	68%[*]	86%[*†]	63%[†]
Postpartum wound infection (cesarean or episiotomy)	75	62%	31%[*]	62%[*]
Postpartum urinary tract infection	63	69%	69%	49%
Any of the above	337	83%[*]	81%	73%[*]

[*] The two percentages in the same row that are marked with this symbol are significantly different by pairwise comparison, p<0.05.
[†] The two percentages in the same row that are marked with this symbol are significantly different by pairwise comparison, p<0.05.

Appendix Table 1: Population-weighted logistic regression model for postpartum maternal readmission, based on ICD-9-CM coded diagnoses originally reported to OSHPD
Vaginal delivery cases

Variable	Parameter Estimate	p value	Odds Ratio	Lower CI for Odds Ratio	Upper CI for Odds Ratio
Intercept	-8.9466	<.0001			
Race: Black	0.5910	<.0001	1.806		
Race: Hispanic	0.5048	<.0001	1.657	1.521	1.805
Race: Other	0.1119	0.0637	1.118	0.994	1.259
Insurance: HMO/Private	-0.0105	0.8022	0.990	0.912	1.074
Insurance: None	-0.4743	0.0005	0.622	0.476	0.814
Drug abuse	0.9544	<.0001	2.597	2.093	3.223
Age: Years	0.2799	<.0001	1.323	1.254	1.396
Age: Years squared	-0.00566	<.0001	0.994	0.993	0.995
Number of antepartum hospitalizations (0-7)	0.6748	<.0001	1.964	1.828	2.110
Chorioamnionitis	0.8614	<.0001	2.366	1.852	3.024
Anemia, specified causes	1.7706	<.0001	5.874	5.193	6.644
Obesity	-0.2490	0.6218	0.780	0.290	2.096
Preeclampsia, mild	1.3866	<.0001	4.001	3.217	4.977
Preeclampsia, severe	-2.2320	0.0003	0.107	0.032	0.360
Diabetes	0.8815	<.0001	2.415	2.101	2.775
Premature labor	-0.4059	<.0001	0.666	0.554	0.802
Previous cesarean	1.5191	<.0001	4.568	4.072	5.124

Appendix Table 2: Population-weighted logistic regression model for postpartum maternal readmission, based on ICD-9-CM coded diagnoses originally reported to OSHPD
Cesarean delivery cases

Variable	Parameter Estimate	p value	Odds Ratio	Lower CI for Odds Ratio	Upper CI for Odds Ratio
Intercept	-4.8020	<.0001			
Race: Black	0.8793	<.0001	2.409	2.039	2.846
Race: Hispanic	-0.4126	<.0001	0.662	0.577	0.760
Race: Other	0.8735	<.0001	2.395	2.075	2.766
Insurance: HMO/Private	-0.4865	<.0001	0.615	0.547	0.691
Insurance: None	-1.7558	<.0001	0.173	0.104	0.287
Drug abuse	0.9759	<.0001	2.654	1.759	4.003
Age: Years	0.0421	0.2434	1.043	0.972	1.119
Age: Years squared	-0.00087	0.1580	0.999	0.998	1.000
Number of antepartum hospitalizations (0-7)	0.0792	0.1614	1.082	0.969	1.209
Anemia, specified causes	0.3133	0.0029	1.368	1.113	1.682
Obesity	1.7503	<.0001	5.756	4.427	7.484
Preeclampsia, mild	-0.0335	0.8178	0.967	0.727	1.286
Preeclampsia, severe	1.5992	<.0001	4.949	3.946	6.206
Diabetes, abnormal glucose tolerance	0.7841	<.0001	2.190	1.831	2.620
Diabetes, other	0.3951	<.0001	1.485	1.058	2.082
Intrauterine death	0.7679	0.0363	2.155	1.050	4.423
Premature labor	-0.0321	0.7327	0.968	0.806	1.164
Previous cesarean	-0.6684	<.0001	0.513	0.456	0.576

Appendix Table 3: Population-weighted logistic regression model for postpartum maternal readmission, based on ICD-9-CM coded diagnoses independently recoded by study staff, plus additional clinical variables abstracted from medical records
Vaginal delivery cases

Variable	Parameter Estimate	p value	Odds Ratio	Lower CI for Odds Ratio	Upper CI for Odds Ratio
Intercept	-6.8512	<.0001			
Race: Black (recoded)	0.4956	<.0001	1.641	1.436	1.876
Race: Hispanic (recoded)	0.1977	<.0001	1.219	1.112	1.336
Race: Other (recoded)	0.2199	0.0006	1.246	1.098	1.413
Insurance: HMO/Private	0.1910	<.0001	1.210	1.114	1.315
Insurance: None	-0.4958	0.0008	0.609	0.456	0.814
Drug abuse (recoded)	1.0379	<.0001	2.823	2.319	3.437
Age: Years (recoded)	0.1036	<.0001	1.109	1.053	1.169
Age: Years squared (recoded)	-0.00271	<.0001	0.997	0.996	0.998
Number of antepartum hospitalizations (0-7)	0.5330	<.0001	1.704	1.563	1.858
Chorioamnionitis (recoded)	0.8622	<.0001	2.368	1.880	2.983
Anemia, specified causes (recoded)	1.1480	<.0001	3.152	2.758	3.602
Obesity (recoded)	0.4516	0.0005	1.571	1.220	2.023
Previous cesarean (abstracted)	0.7638	<.0001	2.147	1.933	2.384
Multiple gestation (abstracted)	0.7033	<.0001	2.020	1.439	2.836
Prior vaginal deliveries (>3, abstracted)	0.7485	<.0001	2.114	1.775	2.518
Obesity (BMI>42.1, abstracted)	1.2536	<.0001	3.503	2.803	4.377
Macrosomia (birthweight >4000 g, abstracted)	1.0728	<.0001	2.924	2.638	3.240
Diabetes requiring insulin (abstracted)	1.1836	<.0001	3.266	2.749	3.881
Antepartum fever (T>38.4, abstracted)	1.1352	<.0001	3.112	2.480	3.905
Psychiatric disease (abstracted)	-0.9291	<.0001	0.395	0.295	0.528
Premature, gestational age <28 weeks (abstracted)	-1.3711	<.0001	0.254	0.174	0.370
Premature, gestational age 28-32 weeks (abstracted)	1.7312	<.0001	5.647	3.912	8.152
Premature, gestational age 32-37 weeks (abstracted)	-0.2890	0.0009	0.749	0.632	0.888
Preeclampsia or eclampsia (abstracted)	1.4252	<.0001	4.159	3.527	4.903
Thick meconium (abstracted)	1.0201	<.0001	2.773	2.220	3.465
Overweight (42.1>=BMI>33.7, abstracted)	0.4142	<.0001	1.513	1.359	1.684
Hemorrhage, 3rd trimester (abstracted)	1.3412	<.0001	3.824	2.974	4.915
Anemia at admission (HCT<33%, abstracted)	0.6966	<.0001	2.007	1.815	2.220

Appendix Table 4: Population-weighted logistic regression model for postpartum maternal readmission, based on ICD-9-CM coded diagnoses independently recoded by study staff, plus additional clinical variables abstracted from medical records
Cesarean delivery cases

Variable	Parameter Estimate	p value	Odds Ratio	Lower CI for Odds Ratio	Upper CI for Odds Ratio
Intercept	-3.3186	<.0001			
Race: Black (recoded)	0.8384	<.0001	2.313	1.931	2.769
Race: Hispanic (recoded)	-0.5176	<.0001	0.596	0.512	0.694
Race: Other (recoded)	1.2449	<.0001	3.472	2.989	4.035
Insurance: HMO/Private	-0.2332	0.0007	0.792	0.692	0.907
Insurance: None	-1.5530	<.0001	0.212	0.127	0.353
Drug abuse (recoded)	1.8278	<.0001	6.220	4.824	8.021
Age: Years (recoded)	-0.1101	0.0035	0.896	0.832	0.965
Age: Years squared (recoded)	0.00179	0.0052	1.002	1.001	1.003
Anemia, specified causes (recoded)	0.5680	<.0001	1.765	1.506	2.068
Obesity (recoded)	0.3662	0.0061	1.442	1.110	1.874
Stillbirth (recoded)	0.4105	0.2830	1.508	0.713	3.189
Previous cesarean (abstracted)	-0.7914	<.0001	0.453	0.398	0.516
Multiple gestation (abstracted)	1.5093	<.0001	4.524	3.414	5.995
Prior vaginal deliveries (>3, abstracted)	0.3042	0.0679	1.355	0.978	1.879
Obesity (BMI>42.1, abstracted)	0.5539	<.0001	1.740	1.390	2.178
Macrosomia (birthweight >4000 g, abstracted)	0.2241	0.0016	1.251	1.089	1.438
Diabetes requiring insulin (abstracted)	2.4882	<.0001	12.040	9.046	16.025
Antepartum fever (T>38.4, abstracted)	0.1179	0.4961	1.125	0.801	1.580
Psychiatric disease (abstracted)	0.9659	<.0001	2.627	2.031	3.398
Premature, gestational age <28 weeks (abstracted)	1.8320	<.0001	6.246	4.870	8.012
Premature, gestational age 28-32 weeks (abstracted)	-1.2848	<.0001	0.277	0.175	0.439
Premature, gestational age 32-37 weeks (abstracted)	-1.3982	<.0001	0.247	0.186	0.328
Preeclampsia or eclampsia (abstracted)	1.2786	<.0001	3.591	3.083	4.184
Thick meconium (abstracted)	0.7552	<.0001	2.128	1.714	2.642
Crash cesarean delivery (abstracted)	0.4432	<.0001	1.558	1.368	1.773

REFERENCES

1. Kazandjian VA, Matthes N, Wicker KG. Are performance indicators generic? The international experience of the Quality Indicator Project. *J Eval Clin Pract* 2003;9:265-276.
2. DesHarnais SI, Forthman MT, Homa-Lowry JM, et al. Risk-adjusted clinical quality indicators: indices for measuring and monitoring rates of mortality, complications, and readmissions. *Qual Manag Health Care* 2000;9:14-22.
3. Ashton CM, Wray NP. A conceptual framework for the study of early readmission as an indicator of quality of care. *Soc Sci Med* 1996;43:1533-1541.
4. Thomas JW. Does risk-adjusted readmission rate provide valid information on hospital quality? *Inquiry* 1996;33:258-270.
5. Rich MW, Beckham V, Wittenberg C, et al. A multidisciplinary intervention to prevent the readmission of elderly patients with congestive heart failure. *N Engl J Med* 1995;333:1190-1195.
6. Mayo PH, Richman J, Harris HW. Results of a program to reduce admissions for adult asthma. *Ann Intern Med* 1990;112:864-871.
7. Madge P, McColl J, Paton J. Impact of a nurse-led home management training programme in children admitted to hospital with acute asthma: a randomized controlled study. *Thorax* 1997;52:223-228.
8. Stewart S, Pearson S, Horowitz JD. Effects of a home-based intervention among patients with congestive heart failure discharged from acute hospital care. *Arch Intern Med* 1998;158:1067-1072.
9. Benbassat J, Taragin M. Hospital readmissions as a measure of quality of health care: advantages and limitations. *Arch Intern Med* 2000;160:1074-1081.
10. Ashton CM, Del Junco DJ, Souchek J, et al. The association between the quality of inpatient care and early readmission: a meta-analysis of the evidence. *Med Care* 1997;35:1044-1059.
11. Melnikow J, Romano PS, Gilbert WM, et al. Vaginal birth after cesarean - in California. *Obstet Gynecol* 2001;98:421-426.
12. Donabedian A. *Introduction to Quality Assurance in Health Care.* Bashshur R, ed. New York: Oxford University Press; 2003.
13. Yasmeen S, Romano PS, Schembri ME, et al. Accuracy of obstetric diagnoses and procedures in hospital discharge data. *Am J Obstet Gynecol* 2006;194(4):992-1001. (See also Section 2 of this report.)
14. Brown S, Small R, Faber B, et al. Early postnatal discharge from hospital for healthy mothers and term infants. *The Cochrane Database of Systematic Reviews* 2002, issue 3. Art. No.: CD002958.
15. Welsh C, Ludwig-Beymer P. Shortened lengths of stay: ensuring continuity of care for mothers and babies. *Lippincotts Prim Care Pract* 1998;2(3):284-291.
16. Bossert R, Rayburn WF, Stanley JR, et al. Early postpartum discharge at a university hospital. Outcome analysis. *J Reprod Med* 2001;46:39-43.
17. Rhodes MK. Early discharge of mothers and infants following vaginal childbirth at the United States Air Force Academy: A three-year study. *Mil Med* 1994;159:227-230.
18. Thompson AH, Alibhai A, Saunders LD, et al. Post-maternity outcomes following health care reform in Alberta: 1992-1996. *Can J Public Health* 2003; 94:104-108.
19. Lieu TA, Wikler C, Capra AM, et al. Clinical outcomes and maternal perceptions of an updated model of perinatal care. *Pediatrics* 1998;102:1437-1444.

20. Tai-Seale M, LoSasso AT, Freund DA, et al. The long-term effects of Medicaid managed care on obstetrics care in three California counties. *Health Serv Res* 2001;36:751-771.
21. Liu S, Heaman M, Kramer MS, et al. Maternal Health Study Group of the Canadian Perinatal Surveillance System. Length of hospital stay, obstetric conditions at childbirth, and maternal readmission: a population-based cohort study. *Am J Obstet Gynecol* 2002;187:681-687.
22. Bashiri A, Smolin A, Sheiner E, et al. Maternal rehospitalization after singleton term vaginal delivery. *J Matern Fetal Neonatal Med* 2003;14:344-348.
23. Hofer TP, Hayward RA. Can early readmission rates accurately detect poor-quality hospitals? *Med Care* 1995;33:234-245.