Evaluating Obstetrical Residency Programs Using Patient Outcomes

David A. Asch, MD, MBA

Sean Nicholson, PhD

Sindhu Srinivas, MD, MSCE

Jeph Herrin, PhD

Andrew J. Epstein, PhD, MPP

Many physicians and non-physicians likely assume that some residency programs tend to produce better physicians than others—either because those residency programs train physicians better or because those residency programs can recruit more capable trainees. Although plausible, these intuitions have not been empirically tested. This information could be useful in at least 2 different ways.[1] First, identifying which training programs produce better physicians and separating out the effects that are due to the ability to attract better trainees might indicate what makes better programs better. Some of these factors might be exportable to other programs, raising the quality of medical education more broadly. Second, by identifying which training programs produce better physicians, patients could use this information when selecting a physician, much as patients in some surgical settings use information on clinician volume when selecting a surgeon and a hospital.[2] Some patients might already be preferentially seeking physicians who have graduated from programs they believe to be elite, but without the evidence to support their intuition.

This study tested the concept that residency programs matter by exploring whether obstetrics and gynecology (OB) residency programs can be evaluated according to the outcomes of the women delivered by the graduates

Context Patient outcomes have been used to assess the performance of hospitals and physicians; in contrast, residency programs have been compared based on nonclinical measures.

Objective To assess whether obstetrics and gynecology residency programs can be evaluated by the quality of care their alumni deliver.

Design, Setting, and Patients A retrospective analysis of all Florida and New York obstetrical hospital discharges between 1992 and 2007, representing 4 906 169 deliveries performed by 4124 obstetricians from 107 US residency programs.

Main Outcome Measures Nine measures of maternal complications from vaginal and cesarean births reflecting laceration, hemorrhage, and all other complications after vaginal delivery; hemorrhage, infection, and all other complications after cesarean delivery; and composites for vaginal and cesarean deliveries and for all deliveries regardless of mode.

Results Obstetricians' residency program was associated with substantial variation in maternal complication rates. Women treated by obstetricians trained in residency programs in the bottom quintile for risk-standardized major maternal complication rates had an adjusted complication rate of 13.6%, approximately one-third higher than the 10.3% adjusted rate for women treated by obstetricians from programs in the top quintile (absolute difference, 3.3%; 95% confidence interval, 2.8%-3.8%). The rankings of residency programs based on each of the 9 measures were similar. Adjustment for medical licensure examination scores did not substantially alter the program ranking.

Conclusions Obstetrics and gynecology training programs can be ranked by the maternal complication rates of their graduates' patients. These rankings are stable across individual types of complications and are not associated with residents' licensing examination scores.

JAMA. 2009;302(12):1277-1283   www.jama.com

of those programs. The advantages of using obstetrics to evaluate the connection between training and clinical outcomes include (1) more than 4 million women giving birth annually in the United States,[3] making delivery one of the most common reasons for hospital care; (2) most women who deliver are healthy, so only limited severity adjustment is needed in evaluating clinical outcomes; and (3) in most cases vaginal deliveries are performed by a single physician and cesarean deliveries are led by a single physician. Furthermore, maternal complications of vaginal and cesarean deliveries, such as hemorrhage, infection, and laceration, occur with sufficient frequency and have enough clinical meaning to patients to serve as markers of quality in maternal care. Risk-adjusted rates of these complications were evaluated as mea-

Author Affiliations: Center for Health Equity Research and Promotion, Philadelphia Veterans Affairs Medical Center, Philadelphia, Pennsylvania (Dr Asch); Leonard Davis Institute of Health Economics (Drs Asch, Nicholson, Srinivas, and Epstein) and Department of Obstetrics and Gynecology (Dr Srinivas), University of Pennsylvania, Philadelphia; Cornell University, Ithaca, New York (Dr Nicholson); and Yale University, New Haven, Connecticut (Drs Herrin and Epstein).

Corresponding Author: David A. Asch, MD, MBA, Leonard Davis Institute of Health Economics, University of Pennsylvania, 3641 Locust Walk, Philadelphia, PA 19104-6218 (asch@wharton.upenn.edu).

sures to judge the quality of care delivered by the graduates of US obstetrical residency programs.

METHODS
Patients, Physicians, and Residency Programs

We examined Florida and New York hospital discharge data between 1992 and 2007, representing every delivery at all nonfederal acute care hospitals in these states. These states were selected because their data contain identifiers for hospitals and physicians, in addition to information on primary and secondary diagnoses, demographic characteristics, procedure use, length of stay, payer, total charges, and admission and discharge status. Cesarean deliveries were identified with an *International Classification of Diseases, Ninth Revision, Clinical Modification* (*ICD-9-CM*) procedure code of 74 in any procedure code field; vaginal deliveries were identified by *International Classification of Diseases, Ninth Revision* (*ICD-9*) diagnosis codes of 650 or 640.0x through 676.9x (where x is 1 or 2) in the principal diagnosis field and the absence of a code for cesarean delivery. The discharge data were augmented with information on the extent of each hospital's OB residency training (including having an OB residency program or hosting an OB rotation) from the National Residency Match Program and information on each physician's sex, specialty, OB residency training program, and residency graduation year from the American Medical Association's Physician Masterfile.

There were 7 130 457 deliveries in Florida and New York during the 16-year period. If there was not high confidence that the delivering physician was identified accurately, deliveries were excluded. Criteria for exclusion included missing or invalid state license number (n=329 052), failing to match the delivering physician to the American Medical Association's Physician Masterfile (n=393 615), the delivering physician performing fewer than 100 deliveries during the entire

study period (to limit the study to obstetricians actively performing deliveries) (n=22 624), failing to find that the delivering physicians reported a primary or secondary specialty of obstetrics, gynecology, or both (n=182 220), failing to find that the delivering physician completing an OB residency (n=33 053), and the delivering physician completing OB residency training after first appearing in the discharge data (n=76 811). After these exclusions, 6 093 082 deliveries remained. Analysis was limited to patients of physicians from residency programs for which we could identify at least 10 physicians. An additional 45 deliveries were excluded for missing or extreme age values (outside 11-55 years). The final analytic data set comprised 4 906 169 deliveries performed by 4124 physicians from 107 US residency programs. The residency programs were distributed among 22 states and the District of Columbia, and represented 43% of the current 249 accredited US OB residency programs.

Maternal Outcomes

The principal study outcome was a binary indicator for maternal complication measured at the patient level (diagnosis codes used to identify maternal complications are shown in eTable 1; available at http://www.jama.com). Maternal complications were analyzed separately by delivery mode. For vaginal deliveries, we measured (1) laceration, (2) hemorrhage, and (3) all others (eg, infectious and thrombotic complications); for cesarean deliveries, we measured (4) hemorrhage, (5) infection, and (6) all others (eg, operative and thrombotic complications). We also measured a composite for each delivery mode (7) and (8), and an overall composite (9) of the 6 individual measures, reflecting any of these maternal complications.

Statistical Analysis

Because the data for this study are naturally nested, with multiple patients associated with each physician and multiple physicians associated with each

residency program, hierarchical generalized linear models (HGLMs) with a logit link function[4-6] were used to assess the independent association between residency program and the 9 maternal complication measures. Patient and physician characteristics were selected for inclusion in the model specification based on a review of the prior literature and clinical judgment. At the patient level, we controlled for demographics (age, racial/ethnic minority status), having Medicaid or no insurance, weekend admission, 34 maternal comorbidities (including prior cesarean delivery, fetal malpresentation, severe hypertension, multiple gestation, antepartum bleeding, herpes, macrosomia, unengaged head, maternal soft tissue disorder, preterm labor, congenital anomalies, oligohydramnios, and polyhydramnios),[7] whether the hospital had an OB residency program or hosted OB residency rotations, and year of hospital discharge.

Patient racial/ethnic status was derived from information in the hospital discharge abstracts as coded at each hospital and included in the state hospital discharge data, consistent with state-specific guidelines for coding race and ethnicity. We included patient racial/ethnic status to control for potential variation in complications that should not be attributable to physicians or residency programs. Because the hospital discharge data lacked unique patient identifiers, we were unable to account for multiple discharges per patient and therefore treated each discharge as an independent observation.

At the physician level, we controlled for physician state (Florida or New York), sex, and years of experience after completing an OB residency. To avoid collinearity with year of discharge, physician experience was measured as of 2007, categorized into quintiles. At the patient level, a random intercept with a normal distribution over physicians was specified and, at the physician level, a random intercept with a normal distribution over the 107 OB residency programs repre-

sented in the analytic data set was specified. For each model, the C statistic was calculated as a measure of its discriminative power.[8] The proportion of variance explained for each model was also calculated.[9]

A risk-standardized complication rate (RSCR) was calculated from the results of the HGLMs for each residency program for each of the 9 complication measures. The RSCR reflects the risk-adjusted program-specific complication rate divided by the estimate of the expected complication rate of the mean residency program (eAppendix; available at http://www.jama.com).[10-12]

Residency programs were ranked for the 3 composite outcomes and each of the 6 individual complication outcomes using their RSCRs. The 6 individual program rankings were compared on a pairwise basis with Spearman rank correlations corrected for multiple comparisons using the Sidak method.[13]

Measures of Clinical Importance

We estimated the adjusted rate of each outcome for each residency program, as well as how much a woman could expect to benefit from being treated by a physician from a high-ranking residency program compared with a low-ranking program. For a woman with the mean values of the patient and physician covariates, we calculated the adjusted rate of a complication assuming she were treated by an average physician trained at each program. We calculated the mean adjusted rate and 95% confidence intervals (CIs) for these rates over all programs in each quintile, and for the difference between top and bottom quintiles, as a measure of absolute risk reduction.

Residency Selection Adjustment

A secondary analysis explored whether the estimated program rankings result from differences in a residency program's ability to attract talented residents vs its ability to improve the residents' skills. Data on medical licensure test scores were obtained from the National Board of

Medical Examiners and the Federation of State Medical Boards. These tests are typically administered near the start of residency. To the extent that licensure examinations are an indicator of underlying trainee ability, adjusting for these test scores could potentially separate contributions of selection and training effects to residency program quality.

The sample included results from 4 distinct tests: 2 versions of the Federation Licensing Examination; the National Board of Medical Examiners Part I, Part II, and Part III; and the US Medical Licensing Examination Step 1, Step 2, and Step 3. These are the 4 major examinations that have been administered to medical school students during the past 50 years; nearly all physicians in our sample (96.8%) took only 1 of these tests. For physicians with multiple scores for the same test, we used the most recent scores. For comparability across test years and versions, we aggregated the test components into a basic science score and a clinical science score and computed version-specific Z scores.

Complete test data were available for 3050 physicians (74.0%) in our analytic sample, representing all 107 residency programs and 3 862 144 deliveries (78.7%). Compared with all physicians, those with complete test data had fewer years of experience (17.3 vs 19.4 years, $P < .001$), were more frequently women (41.9% vs 38.6%, $P = .004$), and were located equally in New York (69.5% vs 69.8%, $P = .74$).

Using this subset of deliveries, the primary analysis was repeated with and without the 2 test Z scores. A Wilcoxon signed rank test was used to compare the distribution of rankings calculated with and without adjustment for test scores. The absolute risk reduction from moving from physicians from bottom-quintile programs to physicians from top-quintile programs with and without adjustment for licensing scores was estimated. The difference in the absolute risk reduction may reflect the clinical effect of attracting more talented residents.

Analyses were performed by using SAS version 9.1.3 (SAS Institute, Cary, North Carolina), Stata version 10.1SE (StataCorp LP, College Station, Texas), and HLM version 6 (Scientific Software International, Lincolnwood, Illinois) software packages. The study was exempted from review by the institutional review boards at the University of Pennsylvania and Cornell University, and the human investigation committee at Yale University.

RESULTS

TABLE 1 shows the patient characteristics and **TABLE 2** shows the maternal comorbidities of the 4 906 169 deliveries in the sample. The crude rate of any major maternal complication among all deliveries was 12.5%. eTable 2 (available at http://www.jama.com) compares the patient characteristics of the sample deliveries with the 2 224 188 deliveries excluded from the analysis. The 2 samples appear very similar, except that excluded patients were disproportionately from Florida, were older, and had Medicaid or no insurance.

eTable 3, eTable 4, and eTable 5 (available at http://www.jama.com) show the estimated coefficients and 95% CIs for the HGLMs used to generate the residency program rankings for the 9 maternal complication outcomes. The model C statistics ranged from 0.646 (any major complication regardless of delivery mode) to 0.775 (infection among cesarean deliveries). The proportion of variance explained by the models ranged between 7.3% (any major complication regardless of delivery mode) and 23.0% (infection among cesarean deliveries). Sample characteristics are presented by quintile of residency program ranking for the outcome of any major complication among all deliveries (**TABLE 3**). Additional information on patient characteristics stratified similarly is shown in eTable 6 (available at http://www.jama.com).

Adjusted rates of complication from physicians trained in the top-quintile programs were substantially lower

Table 1. Patient Characteristics[a]

Characteristics	All Deliveries (N = 4 906 169)	Vaginal Delivery (n = 3 520 989)	Cesarean Delivery (n = 1 385 180)
Maternal complications			
Any major complication	12.5	12.9	11.6
Laceration		6.5	NA
Hemorrhage		3.7	5.8
Infection		NA	4.7
Other		2.8	2.3
Physician characteristics			
New York state	61.9	62.7	60.0
Female obstetrician	29.4	29.6	28.9
Obstetrician years of experience, as of 2007, mean (SD)	20.4 (9.3)	20.5 (9.3)	20.1 (9.3)
Hospital characteristics			
Has OB residency	39.6	40.1	38.1
Hosts OB residency rotations	4.1	4.2	3.8
Maternal demographics			
Age, mean (SD), y	28.1 (6.2)	27.6 (6.1)	29.5 (6.2)
Admitted on Saturday or Sunday	20.9	23.1	15.2
Other race besides non-Hispanic white	44.6	44.7	44.4
Medicaid insurance or uninsured	42.5	44.5	37.5

Abbreviations: NA, not applicable; OB, obstetrics and gynecology.
[a] Data are presented as percentage unless otherwise specified. Other maternal complications are listed in eTable 1 (available at http://www.jama.com).

than from those physicians trained in the bottom-quintile programs (**TABLE 4**). All else equal, a woman choosing an obstetrician who trained at a program in the top tier would face a 10.3% risk of a major complication compared with 13.6% if she chose an obstetrician trained at a program from the bottom tier (absolute difference, 3.3%; 95% CI, 2.8%-3.8%). These differences remained important across the 6 individual complication measures conditional on delivery mode. In general, the bottom-quintile programs had complication rates approximately one-third higher than those of the top-quintile programs.

Consistency Across Complications

The quintile positions of individual residency programs were largely similar regardless of which of the 6 complications for vaginal or cesarean deliveries was used, or whether the programs were judged by the rate of any major complication regardless of delivery mode. The correlation between residency programs' major complication rates in vaginal deliveries (laceration, hemorrhage, other) and its major complication

rates in cesarean deliveries (hemorrhage, infection, other) was 0.51 ($P < .001$). The quintile positions of the 107 residency programs were also similar, whether judged by vaginal or cesarean delivery complication rates. For example, 31% of the 107 residency programs stayed within the same quintile across both measures, 64.5% stayed within adjacent quintiles, and 91% stayed within 2 quintiles. More generally, residency programs that produced physicians with low-adjusted rates of one complication also produced physicians with low-adjusted rates of other complications. For example, **TABLE 5** shows pairwise Spearman rank correlations across residency programs for the individual complications.

Selection Effects

These analyses were repeated for the 74% of obstetricians for whom we had medical licensure scores. When those Z scores were included in the model to reflect differential selection of trainees into programs, the results were largely unchanged. For the complication measures shown in Table 3, the

ranking distributions were statistically no different from each other for the overall composite measure (paired sign rank test, $P = .57$) and each of the 6 individual complication measures (all $P > .32$). Across the 7 outcome measures, the difference between best and worst quintile shrunk by an average of 0.09% in absolute terms (range, 0.02%-0.26%).

COMMENT

Many patients, prospective trainees, medical educators, and those individuals who hire physicians for clinical practices probably share the view that where a physician trained gives at least some indication of how good that physician is currently. This study demonstrates that OB residency programs can be ranked according to the risk-adjusted maternal complications of the women treated by the graduates of those programs, that these rankings are generally consistent across 6 different individual obstetrical complications, and that the expected clinical benefit of moving from treatment by an obstetrician who graduated from a lower-tier program to an obstetrician from a higher-tier program is relatively large (with women in this sample experiencing a 10.3% complication rate when treated by physicians trained in top-quintile programs compared with a rate of 13.6% when treated by physicians trained in bottom-quintile programs). To our knowledge, these findings provide the first empirical support for widely-held intuitions about the clinical implications of variation in medical education. The often large and uniformly positive correlations across the 9 separate measures lend support to the view that rates of individual complications track together at the level of the residency program and suggest that these rates may reflect good measures of overall quality.

These results may have important implications for patients. Little is known about how women choose obstetricians,[14] but it seems unlikely that maternal complication rates currently determine those choices. The informa-

tion is not generally available, use of individual physician-quality information in any setting is limited, and women in less-densely resourced areas or in certain insurance plans may have few local choices. But it is straightforward to determine where an obstetrician trained. If these findings are confirmed and refined, women might select obstetricians in part by where they were trained. The general consistency in programs' rankings despite different measures of quality supports the validity of the measures and also suggests that top programs may be likely to produce physicians who are better in unmeasured ways as well.

These results may also have important implications for medical educators. Stating that one residency program is good or is better than another residency program may mean many different things to different persons, but it should ultimately mean that good programs produce physicians who take care of patients well, and better programs produce physicians who take care of patients better. By that reasoning, judging medical training programs by subsequent patient outcomes places the evaluation of medical training much closer to its purpose than do evaluations based on admission selectivity, board scores, or rankings by news magazines or leaders in the field. A study by Hartz et al[15] found no association between coronary artery bypass graft surgery mortality and the surgeon's training at a collection of schools and residency programs preidentified as prestigious. In our study, the patients' outcomes determined which programs were better than others.

These results also open an opportunity for investigation in other clinical settings. Early work demonstrating that surgical volume is associated with improved outcomes[2,16] led the way for other investigations exploring volume-outcome associations in a wide range of clinical fields—all testing the hypothesis that experience, expressed in the form of annual or cumulative volume, might be related to quality.[17-19]

Physician ability is likely related not just to experience, but also to training and intrinsic aptitude. Aptitude, training, and experience represent a plausible set of individual inputs contributing to physician quality. By that conceptual model, it should not be surprising that training (measured by training site) would also be a determinant of patient outcomes.

We found no evidence for a major selection effect in residency program output. If programs differ substantially in the quality of physicians they graduate, much of that difference might be attributable to the initial quality of the trainees they attract, but we found little difference in effects after adjustment for individual physicians' standardized medical licensure

Table 2. Maternal Comorbidities[a]

Maternal Comorbidities	All Deliveries (N=4 906 169)	Vaginal Delivery (n=3 520 989)	Cesarean Delivery (n=1 385 180)
Previous cesarean delivery	13.6	3.8	38.4
Fetal malpresentation	6.2	1.9	17.3
Antepartum bleeding	1.7	0.7	4.1
Herpes	1.2	0.9	2.1
Eclampsia or severe preeclampsia	1.0	0.4	2.4
Uterine scar unrelated to cesarean delivery	0.2	0.03	0.6
Multiple gestation	1.3	0.6	3.0
Macrosomia	2.4	1.3	5.1
Unengaged fetal head	1.5	0.1	5.2
Maternal soft tissue disorder	2.8	1.2	6.8
Mild preeclampsia or noneclampsia hypertension	5.9	4.8	8.9
Preterm gestation	6.9	5.9	9.5
Congenital fetal CNS anomaly or chromosomal abnormality	0.1	0.1	0.2
Maternal cerebral hemorrhage	0.004	0.002	0.009
Asthma	1.7	1.5	2.2
Maternal renal abnormality	0.1	0.1	0.2
Maternal liver abnormality	0.1	0.1	0.1
Diabetes or abnormal glucose tolerance	0.6	0.5	0.9
Maternal thyroid abnormality	1.3	1.2	1.8
Maternal substance abuse	0.3	0.3	0.3
Maternal mental disorder	1.9	1.8	2.2
Maternal congenital and other heart disease	1.5	1.4	1.8
Isoimmunization	1.8	1.8	1.8
Intrauterine fetal demise	0.05	0.06	0.03
Intrauterine growth restriction	1.3	1.0	2.2
Oligohydramnios	0.6	0.4	1.1
Polyhydramnios	3.1	2.5	4.7
Ruptured membrane >24 h	1.6	1.5	2.0
Chorioamnionitis	1.8	1.1	3.4
Maternal pyrexia or septicemia	1.0	0.9	1.3
Uterine rupture	0.07	0.01	0.2
Maternal hypotension or obstetrical shock	0.1	0.1	0.1
Pulmonary embolism	0.007	0.002	0.02
Obesity	0.5	0.3	1.2

Abbreviation: CNS, central nervous system.
[a] Data are presented as percentage. Maternal comorbidities were based on *International Classification of Diseases, Ninth Revision (ICD-9)* codes developed by Gregory et al.[7]

examination Z scores. This suggests either that these scores do not capture medical students' clinical ability or that skills developed during residency training are more important for producing good maternal outcomes than skills developed during medical school, and residency programs differ in skill development.

Table 3. Sample Characteristics by Quintile Ranking of Residency Programs for Any Major Complication Among All Deliveries (N = 4 906 169)[a]

Characteristics	Residency Program Maternal Complication Rate Quintile				
	High (Q5)	Q4	Q3	Q2	Low (Q1)
Deliveries	(n = 1 217 247)	(n = 930 091)	(n = 973 129)	(n = 703 673)	(n = 1 082 029)
Cesarean	338 784 (27.8)	260 323 (28.0)	287 777 (29.6)	194 909 (27.7)	303 387 (28.0)
Vaginal	878 463 (72.2)	669 768 (72.0)	685 352 (70.4)	508 764 (72.3)	778 642 (72.0)
Physicians	(n = 1030)	(n = 773)	(n = 765)	(n = 633)	(n = 923)
State					
New York	902 (21.9)	597 (14.5)	352 (8.5)	490 (11.9)	539 (13.1)
Florida	128 (3.1)	176 (4.3)	413 (10.0)	143 (3.5)	384 (9.3)
Sex					
Male	647 (15.7)	504 (12.2)	494 (12.0)	343 (8.3)	546 (13.2)
Female	383 (9.3)	269 (6.5)	271 (6.6)	290 (7.0)	377 (9.1)
Years of experience, as of 2007, mean (SD)	20.4 (11.8)	21.1 (11.2)	18.8 (10.7)	18.4 (11.4)	18.0 (10.8)
Residency programs	(n = 22)	(n = 21)	(n = 22)	(n = 21)	(n = 21)
Risk-standardized complication rate, mean (SD [range])	10.9 (0.5 [9.8-11.6])	12.0 (0.2 [11.7-12.2])	12.6 (0.2 [12.3-12.7])	13.0 (0.1 [12.8-13.2])	14.0 (0.9 [13.3-16.2])

[a] Data are presented as No. (%) unless otherwise specified.

Table 4. Adjusted Rates of Maternal Complication by Quintile Ranking of Residency Programs[a]

	Mean Adjusted Rate, % (95% Confidence Interval)					
	Residency Program Maternal Complication Rate Quintile					High-Low Quintile Difference
	High (Q5)	Q4	Q3	Q2	Low (Q1)	
All deliveries, any major complication	10.3 (10.1-10.5)	11.3 (11.3-11.4)	11.9 (11.9-12.0)	12.4 (12.3-12.5)	13.6 (13.1-14.0)	3.3 (2.8-3.8)
Vaginal delivery						
Laceration	5.0 (4.8-5.2)	5.7 (5.6-5.7)	6.2 (6.1-6.2)	6.6 (6.5-6.6)	7.1 (7.0-7.3)	2.2 (1.9-2.4)
Hemorrhage	2.7 (2.6-2.8)	3.1 (3.1-3.1)	3.3 (3.3-3.3)	3.4 (3.4-3.5)	3.8 (3.7-3.9)	1.1 (1.0-1.3)
Other	2.4 (2.4-2.5)	2.8 (2.7-2.8)	3.0 (3.0-3.1)	3.3 (3.3-3.4)	4.2 (3.8-4.5)	1.7 (1.4-2.1)
Cesarean delivery						
Hemorrhage	3.6 (3.5-3.7)	4.2 (4.1-4.2)	4.4 (4.4-4.5)	4.7 (4.7-4.8)	5.5 (5.1-5.8)	1.8 (1.5-2.2)
Infection	3.3 (3.2-3.4)	3.6 (3.6-3.7)	3.8 (3.8-3.9)	4.1 (4.0-4.1)	4.8 (4.5-5.1)	1.5 (1.2-1.8)
Other	1.4 (1.4-1.4)	1.6 (1.6-1.7)	1.8 (1.8-1.8)	1.9 (1.9-1.9)	2.2 (2.1-2.3)	0.8 (0.7-0.9)

[a] Each quintile structure has been created separately for each outcome measure. Other complications of vaginal deliveries include infections and thrombotic complications, and other complications of cesarean deliveries include operative and thrombotic complications.

Table 5. Spearman Rank Correlations of Risk-Standardized Complication Rates Among Residency Programs[a]

	Vaginal Delivery			Cesarean Delivery		
	Laceration	Hemorrhage	Other	Hemorrhage	Infection	Other
Risk-standardized complication rate, mean (SD)	6.5 (0.8)	3.8 (0.4)	3.5 (0.7)	5.8 (0.7)	4.7 (0.6)	2.4 (0.4)
Vaginal delivery						
Laceration	1 [Reference]					
Hemorrhage	0.309[b]	1 [Reference]				
Other	0.352[b]	0.376[b]	1 [Reference]			
Cesarean delivery						
Hemorrhage	0.061	0.691[b]	0.370[b]	1 [Reference]		
Infection	0.037	0.236	0.087	0.081	1 [Reference]	
Other	0.377[b]	0.332[b]	0.531[b]	0.304[b]	0.236	1 [Reference]

[a] Data are Spearman rank correlations unless otherwise indicated. Other complications of vaginal deliveries include infections and thrombotic complications, and other complications of cesarean deliveries include operative and thrombotic complications.
[b] $P < .05$, Sidak corrected for multiple comparisons.[13]

Our study has several limitations. First, we studied deliveries only in New York and Florida. Although large, those states do not represent all residency programs and obstetricians who stay near their training site may systematically differ from those who relocate. Second, because of data limitations, we studied only maternal outcomes and did not include birth outcomes, which may be viewed as more important than maternal outcomes to women choosing their obstetrical care. Third, our risk-adjustment scheme was limited by our use of administrative data; it is possible that important patient characteristics, such as the extent of preterm care, vary across residency program alumni. Fourth, licensure examinations are at best an incomplete measure of the abilities of physicians before their residency education. Fifth, our sample includes obstetricians who completed residency at many different times. A hospital's residency program in the 1960s might differ from its program in the 1990s because of different faculty, the evolution of new clinical techniques that might diffuse across programs differently, or trends in attracting different trainees. Future work can explore how residencies change over time and whether training programs effectively produce different "vintages" of graduates. Finally, although we derived a model for ranking residency programs and those estimates are internally consistent, the concept is new and our derived model should be independently validated before taking actions to select physicians or change educational programs.

Separate from these methodological limitations, stakeholders might object to the interpretation and use of the results. Where a physician trained provides a signal about later quality, but it is a limited signal. C statistics, reflecting model discrimination, ranged from 0.65 to 0.78. The individual quality of physicians may improve or decline, but

they can never change the residency program in which they trained. A patient may be more likely to identify a higher-quality physician by using rather than ignoring residency program rankings; however, because of individual physician variation, she cannot be certain that a particular physician who trained at a top-ranked program will have generally better outcomes. Many measurable characteristics are associated with quality and our study measures only one of them.

Our study also has several strengths. The study reflects nearly 5 million deliveries, more than 4000 obstetricians, and more than 100 residency programs. The results are adjusted for a large number of patient and physician characteristics. The results are robust across different delivery modes and medical complications.

CONCLUSIONS

Where an obstetrician completed residency may provide a meaningful and consistent signal about the risk of maternal complications among that obstetrician's patients. These rankings are stable across individual types of complications and are not associated with residents' licensing examination scores.

REFERENCES

1. Asch DA, Epstein A, Nicholson S. Evaluating medical training programs by the quality of care delivered by their alumni. *JAMA*. 2007;298(9):1049-1051.
2. Birkmeyer JD, Stukel TA, Siewers AE, Goodney PP, Wennberg DE, Lucas FL. Surgeon volume and operative mortality in the United States. *N Engl J Med*. 2003; 349(22):2117-2127.
3. Births: Preliminary Data for 2005. http://www.cdc.gov/nchs/products/pubs/pubd/hestats/prelimbirths05/prelimbirths05.htm. Accessed July 28, 2009.
4. Bryk AS, Raudenbush SW. *Hierarchical Linear Models: Applications and Data Analysis Methods.* London, England: Sage Publications; 1992.
5. Goldstein H. *Multilevel Statistical Models.* London, England: Arnold; 1995.
6. Normand SL, Shahian DM. Statistical and clinical aspects of hospital outcomes profiling. *Stat Sci*. 2007; 22(2):206-226.
7. Gregory KD, Korst LM, Gornbein JA, Platt LD. Using administrative data to identify indications for elective primary cesarean delivery. *Health Serv Res*. 2002; 37(5):1387-1401.
8. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361-387.
9. Snijders T, Bosker R. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling.* London, England: Sage Publications; 2000.
10. Shahian DM, Torchiana DF, Shemin RJ, Rawn JD, Normand SL. Massachusetts cardiac surgery report card: implications of statistical methodology. *Ann Thorac Surg*. 2005;80(6):2106-2113.
11. Hannan EL, Wu C, Ryan TJ, et al. Do hospitals and surgeons with higher coronary artery bypass graft surgery volumes still have lower risk-adjusted mortality rates? *Circulation*. 2003;108(7):795-801.
12. Krumholz HM, Wang Y, Mattera JA, et al. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with heart failure. *Circulation*. 2006;113 (13):1693-1701.
13. Sidak Z. Rectangular confidence regions for the means of multivariate normal distributions. *J Am Statist Assoc*. 1967;62:626-633.
14. Hoerger TJ, Howard LZ. Search behavior and choice of physician in the market for prenatal care. *Med Care*. 1995;33(4):332-349.
15. Hartz AJ, Kuhn EM, Pulido J. Prestige of training programs and experience of bypass surgeons as factors in adjusted patient mortality rates. *Med Care*. 1999; 37(1):93-103.
16. Hannan EL, Racz M, Ryan TJ, et al. Coronary angioplasty volume-outcome relationships for hospitals and cardiologists. *JAMA*. 1997;277(11):892-898.
17. Sheikh K. Utility of provider volume as an indicator of medical care quality and for policy decisions. *Am J Med*. 2001;111(9):712-715.
18. Sheikh K. Reliability of provider volume and outcome associations for healthcare policy. *Med Care*. 2003;41(10):1111-1117.
19. Luft HS. From observing the relationship between volume and outcome to making policy recommendations: comments on Sheikh. *Med Care*. 2003; 41(10):1118-1122.