The Journal of THORACIC AND CARDIOVASCULAR SURGERY

SURGERY FOR ADULT CARDIOVASCULAR DISEASE

VOLUME-OUTCOME RELATIONSHIPS IN CARDIOVASCULAR OPERATIONS: NEW YORK STATE, 1990-1995

Josephine A. Sollano, MPH
Annetine C. Gelijns, PhD
Alan J. Moskowitz, MD
Daniel F. Heitjan, PhD
Suzanne Cullinane, BS,
Ted Saha, BS
Jonathan M. Chen, MD
Patrick J. Roohan, MS
Keith Reemtsma, MD
Eileen P. Shields, BA

Background: It has been known for nearly 20 years that, in cardiovascular operations, a significant inverse relationship exists between clinical outcomes and the volume of procedures performed. Interestingly, this relationship persists 2 decades after it was recognized. **Objective:** The purpose of this study was to examine the relationship between hospital volume and in-hospital deaths in 3 cardiovascular procedures: coronary artery bypass grafting, elective repair of abdominal aortic aneurysms, and repair of congenital cardiac defects. **Methods:** The database includes all patients who were hospitalized in New York State during the years 1990 to 1995. Using standard logistic regression techniques, we analyzed the relationship between hospital volume and outcome. **Results:** No correlation exists between hospital volume and in-hospital deaths in coronary artery bypass grafting. Statewide, 31 hospitals performed 97,137 operations over the 6-year period (overall mortality rate, 2.75%). By contrast, most of the hospitals statewide (195 of 230 hospitals) performed 9847 elective abdominal aortic aneurysm repairs with an overall mortality rate of 5.5%. In abdominal aortic aneurysm operations, a significant inverse relationship between hospital volume and in-hospital deaths was determined. Sixteen hospitals performed 7199 repairs for

From the International Center for Health Outcomes and Innovation Research (InCHOIR), the Department of Surgery, Columbia University, and the New York State Department of Health.

Read at the Seventy-eighth Annual Meeting of The American Association for Thoracic Surgery, Boston, Mass, May 3-6, 1998.

Received for publication May 8, 1998; revisions requested June 30, 1998; revisions received Oct 20, 1998; accepted for publication Nov 3, 1998.

Address for reprints: Josephine A. Sollano, MPH, Assistant Professor of Clinical Surgical Sciences, Columbia University College of Physicians & Surgeons, Director, Office of Service Line Data and Analytical Support, New York Presbyterian Hospital, 635 West 165th St, 5th Floor, Room 503, New York, NY 10032.

J Thorac Cardiovasc Surg 1999;117:419-30

Copyright © 1999 by Mosby, Inc.

0022-5223/99 \$8.00 + 0 12/6/95596

congenital cardiac defects. A significant inverse relationship (which was most pronounced for neonates) was found between volume and death. Conclusions: The importance of these findings lies in the rather striking difference between the volume-outcome relationship found for operations for abdominal aortic aneurysms and congenital cardiac defects and the lack of such a relationship for coronary artery bypass grafting. This observation may be largely explained by the quality improvement program in New York State for bypass operations since 1989. If so, these results have important implications for expanding the scope of quality improvement efforts in New York State. (*J Thorac Cardiovasc Surg* 1999;117:419-30)

The current environment of market forces in health care has renewed interest in exploring the relationship between volume and outcomes to discern variations in performance that are related to case load. Admittedly, managed care organizations and large purchasers up to now have focused mainly on price as a basis for contractual decisions. But as their experience in this market evolves, quality may become an increasingly important consideration for decision-making.

It has been known for nearly 20 years that, in cardiovascular operations, there is a significant inverse relationship between clinical outcomes and the volume of procedures performed by health care providers.¹ This recognition has led to a variety of responses. At one end of the spectrum lies a regulatory response, which has involved the regionalization of specialty care in a few institutions and the closing of low-volume programs. At the other end of the spectrum is the "laissez-faire" approach; the observations about volume-outcome relationships ostensibly led to no change in providers' or purchasers' behavior. The middle ground contains the creation of quality improvement programs, which continuously survey outcomes and give performance data to providers to encourage them to improve care. We believe that the 20-year mark is an appropriate time to assess (1) whether continued volume-outcome relationships exist in cardiovascular surgery, (2) what the response has been if they do exist, and (3) whether these efforts have been adequate in achieving the goal of better patient care.

We examine volume-outcome relationships in 3 categories of cardiovascular procedures (coronary artery bypass grafting [CABG], operations for congenital cardiac disease, and abdominal aortic aneurysm [AAA] repairs) in New York State during the years 1990 through 1995. The reasons for selecting these procedures were 3-fold. First, they reflect very different annual patient volumes. In New York State, the average

annual number of patients undergoing operations during the 6 years of study were 16,190 CABG procedures, 1641 repairs of nonruptured AAAs, and 1199 pediatric cardiac procedures. The large number of patients undergoing CABG clearly flags this as an important public health issue. Although the population of neonates and infants is considerably smaller, the aggregate number of life-years at risk is substantial, given the nearly 8 decades of life expectancy for newborn infants. Second, in all 3 areas, the "red flag" of significant inverse relationships between volume and outcome has been raised. However, these areas have been studied with varying degrees of intensity; CABG operations have received the most attention in the literature, and pediatric cardiac operations have received the least.²⁻⁸ Finally, there has been a major difference in the professional and regulatory approach toward managing quality. In CABG surgery, we have seen the most comprehensive approach, regulatory intervention, and the creation of a continuous quality improvement program. This program has been organized under the auspices of the Cardiac Advisory Committee, in conjunction with the New York State Department of Health. It entails standardized data collection; feedback of risk-adjusted results to surgeons, hospitals, and the public; the distribution of predictive instruments to surgeons (which allow them to make preoperative predictions); and the creation of specific quality improvement efforts at individual hospitals.²⁻⁶ In pediatric cardiac operations, clinical data are routinely collected by the Cardiac Advisory Committee, but until recently, analyses were not given to the providers.⁸ By contrast, there is no statewide effort to monitor the outcomes of AAA operations.

Methods

Data sources. Discharge abstract data for individual patients who underwent any of the 3 types of cardiovascular

Table I. Complexity categories

Category	Operation description	ICD-9-CM procedure codes
1 (n = 918)	Closed valvotomy	35.0-35.04
	Atrial septectomy	35.41-42
	Repair of atrial septal defect	35.50-52, 35.60-61, 35.70-71
	Patent ductus arteriosus closure or pulmonary artery band	38.85
	Systemic-pulmonary artery shunt	39.0
	Thoracic vessel repair, direct	38.35
2 (n = 1069)	Open valvotomy	35.1-35.14
	Annuloplasty	35.34
	Infundibulectomy	35.34
	Subaortic stenosis resection	35.35
	Repair of ventricular septal defect	35.53, 35.62, 35.72
	Repair of endocardial cushion defect	35.54,35.63,35.73
	Repair of tetralogy of Fallot	35.81
	Thoracic vessel repair, prosthesis	38.45
	Septal surgery, unspecified	35.70, 35.39
	Revision of procedure	35.95
3 (n = 563)	Valve replacement	35.2-35.28
	Repair of total anomalous pulmonary venous connection	35.82
	Repair of truncus arteriosus	35.83
	Atrial inversion	35.91
	Right ventricular-pulmonary artery conduit	35.59
	Left ventricular-aorta conduit	35.93
	Cavopulmonary shunt	39.59
	Repair of vessel, including aortopulmonary window	39.59
4 (n = 283)	Arterial switch operation	35.84
	Total cavopulmonary anastomosis	35.94
	Diagnosis of hypoplastic left heart syndrome	746.7 ^a

Note: Search included "Other operations on valves of the heart" (35.99); no cases found.

(Adapted from Jenkins KJ, Newburger JW, Lock JE, Davis RB, Coffman GA, Iezzoni LI. In-hospital mortality for surgical repair of congenital heart defects: preliminary observations of variation by hospital caseload. *Pediatrics* 1995;95:323-30.)

^aICD-9-CM diagnostic code.

operations during the period from 1990 through 1995 were extracted from the Statewide Planning and Research Cooperative System (SPARCS). These data are maintained by the New York Department of Health and contain information on every patient discharged from an acute care facility, including patient disposition (eg, death), age, gender, principal procedure, and principal diagnosis codes. Each patient record also contains a code that identifies the discharging hospital and whether the patient was transferred from another hospital or a long-term care facility.

The International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) was used to define 3 cohorts of patients.

Patients who underwent elective AAA repairs had a principal ICD-9-CM procedure code of 38.44 (resection of abdominal aorta), with a primary or secondary diagnosis code of 44.14 (abdominal aneurysm without mention of rupture). Patients undergoing operations for ruptured AAAs were excluded because emergency operations of this sort have an across-the-board high mortality rate of 50% and do not necessarily reflect the institution's experience with a planned procedure. Because there is no quality improvement program in

place for AAA operations in New York State, there is no incentive for hospitals or surgeons to overuse the emergency category; the exclusion of patients who underwent emergency AAA repair thus should not impose a bias in our analysis.

Patients who underwent CABG had ICD-9-CM procedure codes in the range from 36.10 through 36.19 (broadly defined as bypass anastomosis for heart revascularization) as a principal procedure. This included single and multiple aorta-coronary bypasses, single and double internal thoracic-coronary artery bypasses, and other bypass anastomoses.

Patients with congenital heart disease were selected by primary and secondary ICD-9-CM procedure codes that were derived from a recent article by Jenkins and colleagues.⁷ The procedure codes and the number of patients in each category are given in Table I. We excluded all premature patients or patients with extremely low birth weight (diagnostic codes, 765.0-765.17) and those patients who underwent closure of an isolated patent ductus arteriosus at less than 3 months of age. In these categories of patients, other factors far outweigh the effect of surgical expertise on patient outcome. We also excluded infants whose cardiac or chest procedures were a result of trauma and who did not have a diagnostic code indi-

Table II. Logistic regression results*: New York State, 1990-1995

Procedure	Variables	Coefficient	P value	OR	95% CI	C _{stat}	HL
CABG (n = 97, 137)	Hospital volume/10,000	0.0014	.88	1.001	0.983-1.020	0.669	0.917
	Female	0.3975	.001	1.488	1.374-1.612		
	Age	0.0587	.001	1.060	1.056-1.065		
Nonruptured AAA repair (n = 9847)	Hospital volume/100	-0.2459	.001	0.782	0.722-0.847	0.679	0.677
	Female	0.2224	.02	1.249	1.029-1.517		
	Age	0.0783	.001	1.081	1.068-1.095		
Pediatric congenital cardiac operation: all ages (n = 7199)	Hospital volume/100	-0.0579	.001	0.944	0.921-0.967	0.617	0.034
	SCC I	Reference group					
	SCC II	-0.3191	.007	0.727	0.577-0.916		
	SCC III	0.5037	.001	1.655	1.292-2.120		
	SCC IV	0.9437	.001	2.569	1.915-3.448		

HL, Hosmer-Lemeshow.

*The prediction of in-hospital death. Hospital volume is not a predictor of in-hospital death for CABG but is a significant predictor of outcome for AAA repair and pediatric congenital operations. Age and female sex are significant independent predictors of death for both CABG and AAA operations.

cating a congenital cardiac anomaly. Finally, all atrial, septal, and ventricular defects that were repaired by transcatheter interventions (as evidenced by a lack of procedure code 39.61 [ie, cardiopulmonary bypass]) were eliminated. We made one exception to the classification by including only those procedures from ICD-9 category 39.59, "repair of other vessels," that also coded for cardiopulmonary bypass.

The analysis of pediatric data was stratified by age (up to 30 days, 31 days-1 year, 1-12 years, and 13-17 years) and adjusted for the complexity of the surgical repair (in neonates we also stratified for procedure complexity). The surgical complexity categories (SCC) were derived from Jenkins and colleagues⁷ (Table I) and were subsequently used by Hannan and colleagues⁸ in a more recent article. The least complex surgical repairs are found in SCC I, whereas the most complex repairs are in SCC IV. Hospital volume was scaled into smaller units to accommodate the interpretation of the logistic model.

The regression analysis examined the relationship between hospital volume and in-hospital deaths. Hospital volume was defined as the number of times each type of operation was performed in a facility over the 6-year period and was calculated by aggregating patient data by the hospital identifier code for each of the surgical cohorts. To evaluate whether there was a trend over the 6-year period of study, we re-examined the data by 2-year intervals. We did not have access to physician identifiers and, therefore, do not address physician volume-outcome relationships.

Data analysis. The relationship of hospital volume to inpatient death was examined with the use of a standard logistic regression procedure. The dependent variable was a binary indicator of in-hospital death. For CABG and AAA repair, we calculated the probability of dying for each patient, after controlling for differences in statewide hospital volume, age, and sex in the baseline-regression model. The baseline pediatric models were stratified by age and controlled for the complexity of procedure but not controlled for sex (in the

neonates, as mentioned, we also stratified for procedure complexity). In addition, we undertook separate models for each level of surgical complexity within the neonate category when significant interaction between surgical complexity level and hospital volume was detected. For each model, we explored the effect of controlling for a variety of other independent variables, including race, payer status, admission status, transfer in from another acute health care facility, and service intensity weights. These weights were used as a proxy for illness severity and are based on diagnosis-related group (DRG) ratings of cost for care. We express the relationship between the outcome and predictor variables as odds ratios (ORs) and represent uncertainty using 95% confidence intervals (CIs). An OR of less than 1.0 denotes the existence of an inverse volume-outcome relationship.

Model performance was measured by the static compliance (C_{stat}), which calculates how well the logistical model discriminates between those patients who lived and those patients who died; a value of 1 indicates perfect discrimination. Additionally, we performed a goodness-of-fit test for each logistic model with the Hosmer-Lemeshow statistic. The Hosmer-Lemeshow statistic compares the observed number of individuals with each outcome with the number expected, as derived by the logistic equation.¹⁰ Large Hosmer-Lemeshow values indicated that the observed number of individuals is different from the expected number of individuals, thereby providing evidence of a poor fit or, more simply, a logistic model with poor predictive abilities.

Hospitals were grouped by volume; each group had an equivalent number of patients, at least 6 deaths, and equal intervals of volume between ranges. The smallest cohort was grouped into 10 volume ranges, and the largest cohort was grouped into 15 volume ranges. For each volume range, we computed the mean actual mortality rate and mean expected mortality rate, which were based on probabilities predicted by the baseline regression analysis.

Fig 1. Actual versus expected in-hospital mortality rates for CABG operations in New York State, 1990-1995. The average hospital volume per annum is shown for a 6-year period. The number of hospitals in each category appears at the top of each solid bar. No volume-outcome relationship is noted.

Fig 2. Actual versus expected in-hospital mortality rates for nonruptured AAA repair in New York State, 1990-1995. The average hospital volume per annum is shown for a 6-year period. The number of hospitals in each category appears at the top of each solid bar. A significant inverse volume-outcome relationship is noted.

Results

The total 6-year cohort for CABG comprised 97,137 patients and for elective AAA operation comprised 9847 patients. The 6-year cohort for pediatric congenital cardiac operation comprised 7199 patients overall, including 1045 neonates (up to 31 days of age), 1985 infants (31 days–1 year of age), 3735 children (1-12

years of age), and 434 adolescents (13-17 years of age).

CABG. In New York State, where certificate-of-need regulation is in place as a mechanism for planning the distribution of heart surgery programs, 31 hospitals are presently certified to perform CABG. A total of 31 hospitals performed on average 3703 procedures over the

Fig 3. Hospital volume is shown for a 6-year period for operations for congenital cardiac defects in neonates (up to 30 days of age) in New York State, 1990-1995. The number of hospitals in each category appears at the top of each bar.

Fig 4. Hospital volume is shown for a 6-year period for operations for congenital cardiac defects in infants (31 days–1 year of age) in New York State, 1990-1995. The number of hospitals in each category appears at the top of each bar.

6-year period, ranging from 70 cases in the lowest volume hospitals to 1421 cases in the highest volume hospitals, annually. The overall mortality rate for CABG was 2.75% during the study period. In the baseline regression model, we found no significant association between hospital case volume and death for patients who underwent CABG (OR, 1.001 for hospital volume/10,000 cases; Table II). We plotted the observed

and expected mortality rates in Fig 1. Examining the same data set by 2-year intervals did not reveal an association between in-hospital deaths and hospital volume. We then expanded the model by 9 variables to include other patient demographics, insurance status, service intensity weights, and mode of admission (ie, elective, urgent, emergency, transferred from another acute care facility). The association between hospital volume and

Table III. Pediatric congenital cardiac operations*: New York State, 1990-1995

Procedure	Variables	Coefficient	P value	OR	95% CI	C _{stat}	HL
Neonates	Hospital volume/100						
(1-30 days	(baseline model)	-0.4520	.001	0.636	0.546-0.742	0.64	0.05
of age; n = 1045)	SCC I (n = 455)	-0.1482	.18	0.862	0.695-1.069	0.53	0.11
	SCC II (n = 167)	-0.5640	.01	0.569	0.361-0.895	0.64	0.21
	SCC III (n = 201)	-0.5197	.001	0.595	0.443-0.799	0.68	0.13
	SCC IV (n = 222; individual models)	-1.2204	.001	0.295	0.167-0.522	0.77	0.36
Infant	Hospital volume/100	-0.3288	.001	0.720	0.622-0.832	0.62	0.21
(31 days-1 year	SCC I (Reference group)						
of age; n = 1985)	SCC II	-0.3034	.13	0.738	0.500-1.091		
	SCC III	0.5079	.03	1.662	1.048-2.636		
	SCC IV	0.9258	.03	2.524	1.082-5.888		
Childhood	Hospital volume/100	-0.1199	.06	0.887	0.784-1.004	0.72	0.33
(1-12 years	SCC I (Reference group)						
of age; n = 3735)	SCC II	1.1591	.001	3.187	1.557-6.523		
	SCC III	1.8703	.001	6.490	3.091-13.630		
	SCC IV	3.0080	.001	20.247	9.644-42.509		
Adolescence	Hospital volume/100	1.0946	.60	2.988	0.052-173.300	0.70	0.75
(13-17 years	SCC I (Reference group)						
of age; n = 434)	SCC II	-0.2824	.78	0.754	0.104-5.447		
	SCC III	1.1639	.15	3.202	0.651-15.760		
	SCC IV	-10.760	.98	0.000	0.000-999.00		

HL, Hosmer-Lemeshow.

*Hospital volume, age, and surgical complexity as predictors of in-hospital death in pediatric patients.

deaths in the expanded model remained the same. In both models, age and female sex were significant predictors of death.

AAA repairs. During the study period, there were 195 hospitals in New York State that performed elective AAA repair. The overall mortality rate for AAA repair was 5.5% over the 6 years. On average, the lowest volume hospitals performed only 2 cases on a yearly basis (with an average mortality rate of 13%), whereas the highest volume center performed 27 cases per year on average with an overall mortality rate of less than 2%. The baseline regression model revealed a significant inverse relationship between hospital case volume and in-hospital deaths (OR, 0.782 for hospital volume/100 cases; 95% CI, 0.722,0.847; $P < .001$; Table II). Using the results of this model, we plotted the observed and expected mortality rates for the different volume groups in Fig 2. Reanalyzing the data in 2-year intervals continued to show the association between low volume and adverse outcome (1990-1991: OR, 0.419; 95% CI, 0.278-0.631; 1992-1993: OR, 0.562; 95% CI, 0.378,0.835; 1994-1995: OR, 0.496; 95% CI, 0.323,0.764). Expanding the model by the 9 variables yielded a similar association between hospital volume and deaths (OR, 0.827; 95% CI, 0.756,0.904; $P < .001$). As with CABG, age and female sex were significant predictors of death for elective repair of AAAs (Table II).

Pediatric cardiac surgery. Sixteen hospitals have a certificate of need for performing surgery of congenital heart disease in New York State. For all pediatric patients who undergo heart procedures, the overall mortality rate for the entire cohort was 7.3% (<31 days of age, 19.8%; 31 days-1 year of age, 9%; 1-12 years of age, 3.5%; 13-20 years of age, 2.5%). A significant inverse relationship was demonstrated between volume and death in all pediatric patients (OR, 0.944 for hospital volume/100 cases; 95% CI, 0.92, 0.97; $P < .001$; Table II).

The relationship was most pronounced for neonates (<31 days of age) in hospitals with an annual volume below 50 patients that displayed mortality rates greater than twice those observed in the highest volume hospital. Fig 3 depicts the observed and expected deaths by hospital volume. For patients aged 31 days to 1 year, a significant inverse relationship was also found (Fig 4). A marginal inverse volume-outcome relationship was found ($P = .579$) in patients aged 1 to 12 years. No significant relationship was found in patients aged 13 to 17 years. Although the overall effect of volume to outcome exists in 3 of the 4 age groups, the magnitude of the effect diminishes with increasing age.

We then examined the volume-outcome relationship for the entire pediatric population, adjusting for surgical complexity. The interaction between surgical complexity and volume was most pronounced for neonates

(Table III). As a result, we reanalyzed the neonatal group, stratifying by surgical complexity. We found no volume-outcome relationship among neonates in SCC I (OR, 0.862 for hospital volume/100 cases); however, significant relationships were found in SCC II, III, and IV (OR, 0.569, 0.595, 0.295 for hospital volume/100 cases; Table III).

Expanding the baseline model in the neonates to include 9 additional variables (excluding Medicare and advanced age and including gender), hospital volume remained a significant predictor. The addition of the 9 variables resulted in findings similar to the baseline model (OR, 0.566; 95% CI, 0.471, 0.681; $P < .001$). Although the fit of the model is marginally better (C_{stat} , 0.67 vs 0.64), we focus on the simpler model because of concerns about the accuracy of reporting, such variables as race or payer status.

For all of the data sets, robust standard errors computed under the generalized estimating equation approach were similar to those from the standard logistic regression, suggesting that correlation within hospitals was negligible.

Discussion

Our analysis reveals that for several, but not all, areas of cardiovascular operations, important variations remain in patient outcomes across hospitals in New York State. The statewide volume of CABG operations in New York State increased from 13,644 in 1990 to 18,821 in 1995 among the 31 hospitals that performed heart operations. Despite an increase in the severity of illness, our data suggest that overall mortality rates declined from 3.28% to 2.49% during the same time period. We did not find a significant inverse relationship between hospital volume and outcomes over the entire period from 1990 through 1995.

After the seminal studies by Luff and colleagues,¹ a considerable amount of literature has emerged concerning CABG, describing the existence of an inverse relationship between the outcomes and the experience of providers. In New York State, Hannan and colleagues²⁻⁶ showed in various studies that there was a marginal relationship between hospital volume and outcomes but a substantial relationship between the volume of activity of the individual surgeon and outcomes. Furthermore, they demonstrated that since 1989 there was a notable decline in risk-adjusted mortality rates for various surgeon volume groups, but the risk-adjusted mortality rates for patients of surgeons who performed 50 or fewer surgical interventions in 1992 was still 36% higher than the mortality rates for patients of surgeons who performed 150 or more CABG procedures.² No studies of volume-outcome

relationships in New York State after 1992 were reported to the Cardiac Advisory Committee's clinical database. Our analysis focused on hospital performance only, and we found no significant volume-outcome relationship in this area between 1990 and 1995, nor did we find any such relationships after disaggregating the data into 2-year periods. The possibility exists that there still is a relationship between surgeon volume and outcome, but if the trend that Harman and colleagues found persists, it is likely to be a relatively small one.

AAA repairs and pediatric cardiac operations, however, present a very different picture. Already in 1989, investigators showed that elective operations for the repair of AAA displayed a significant inverse relationship between volume and death in New York State.⁹ This analysis, and our previous research, suggests that this relationship persists over a decade later,¹¹ with the vast majority of hospitals still performing very few surgical procedures annually (84% of the 195 hospitals that perform AAA repairs perform fewer than 10 elective cases per year).

Our conclusions for the pediatric population are 3-fold. First, pediatric patients undergoing cardiac surgical procedures in general demonstrate a significant inverse relationship between hospital volume and hospital death. Second, when stratified by age, complexity overshadowed volume as a significant determinant of death among neonates, infants, and children up to age 12 years, but patients older than 12 years did not demonstrate these differences. Finally, this relationship is most pronounced in neonates for whom hospitals with volumes below 50 patients per year had mortality rates that were 2 times higher than in the highest hospital volume group. These findings are consistent with those of Jenkins and colleagues,⁷ who reported similar results in a review of congenital heart defect operations in California in 1988 and in Massachusetts in 1989. In these cohorts, Jenkins and colleagues demonstrated a 3-fold increase in in-hospital deaths when comparing the ORs of centers performing 10 to 100 procedures per year with those centers performing more than 300 procedures. A recent study, with the use of the Cardiac Surgery Reporting System database, documents similar results for New York State; both hospital and surgeon volume are significantly and negatively associated with in-hospital deaths.⁸ One difference between the 2 studies is that Hannan and colleagues found the volume-outcome differences to persist for both high-complexity and low-complexity cardiac procedures, whereas in our study the lowest complexity category did not show a significant inverse relationship between adverse outcomes and hospital experience.

Our analysis suggests not only that variations in hospital performance exist within the state but also that there are important differences in the approaches used to improve the quality of care in these 3 clinical areas. In CABG procedures, the Cardiac Advisory Committee and the New York State Department of Health in 1989 created a prospective clinical data collection effort for cardiac operations, patients, and all hospitals that voluntarily report information regarding risk factors, deaths, and complications to this Cardiac Surgery Reporting System. These data are used to analyze actual and risk-adjusted outcomes, and these data are reported to the public, surgeons, and hospitals as a basis for quality improvement efforts. When analyzed by surgeon volume, several studies have attributed part of the decline in overall risk-adjusted deaths to these quality improvement efforts.²⁻⁶ In particular, these studies found that many low-volume surgeons stopped performing CABG in New York State. Not all of the decline, however, could be explained by New York's quality improvement program, because a similar trend was found in states with no such program. Pediatric cardiac operations also fall within the cardiac operations reporting system, but this has, until now, not led to a continuous quality improvement program; data have not been fed back on a regular basis to hospitals, and only recently have the first risk-adjusted mortality rates for various hospital and surgeon volume ranges been published. In AAA repairs, no prospective statewide patient registries exist that can form the foundation for quality improvement efforts.

Why do these differences exist? The CABG operation is a high-volume surgical procedure that was introduced just after certificate-of-need regulation was implemented in New York State in 1966. The existence of certificate-of-need regulation meant that the state and providers were already concerned with the quality of heart operation programs, which then could provide the foundation for establishing a continuous quality improvement program. Pediatric cardiac surgery also falls under certificate-of-need regulation in the state of New York; however, the data collected within the cardiac surgery reporting system were not analyzed and reported to providers on a regular basis, probably because pediatric cardiac surgery is a catch-all for many procedures and involves much lower patient volumes. Hospitals do not need a certificate of need to perform AAA repair, and compared with CABG, the volume of patients for this procedure is much lower. These factors probably contributed to the absence of a statewide monitoring and quality improvement system.

What lessons can we draw from comparing the performance of providers in the area of cardiovascular

surgery in New York State using the SPARCS database? We would underscore that there are serious limitations to the analysis of risk-adjusted outcomes on the basis of hospital discharge databases, such as SPARCS.^{9,11} First, these databases do not contain the physiologic data with which to distinguish patients. Second, for the data that are coded, there is substantial variation in accuracy and completeness, resulting in both overreporting and underreporting. Such errors could be unintentional or deliberate by those who might game the system. Third, the level of detail contained in these databases do not allow one to distinguish pre-existing conditions from complications of the hospitalization. As a result of these limitations, we chose in our baseline analysis to control only for age, sex, and volume, which are less subject to error, and we still found a significant volume-outcome relationship for 2 of the 3 procedures, which remained when additional predictive variables were considered.

Despite the limitations of hospital discharge databases, these data strongly indicate the existence of variations in outcomes among hospitals in the treatment of congenital heart disease and aortic aneurysms. In the area of AAA repairs, such variations have been identified before and should be a strong stimulus for the community of vascular and general surgeons to develop the infrastructure for rigorous clinical data collection as a basis for quality improvement. We believe that this should be a particularly strong stimulus because, otherwise, regulatory and purchasing decisions might be based on data that are not of research quality and are inadequate for judging the severity of illness. The data collection infrastructure can form the basis for a quality improvement program.

In pediatric cardiac surgery, prospective clinical data collection is in place but as yet has not led to a continuous quality improvement program, as it has in adult coronary revascularization. The recent analysis of pediatric data collected by the Cardiac Advisory Committee, and our findings here, offer a strong rationale for building a continuous quality improvement program for pediatric surgery. With respect to highly complex neonatal cardiac procedures, where the volume-outcome relationship was particularly strong, the break between institutions with comparatively high and low mortality rates clearly is in excess of 100 patients, and possibly over 200 patients. Given the volume needed to support low-death operations, the state population could support many fewer institutions than the 16 certified hospitals where this procedure is performed now. This constitutes a strong argument for centralization of heart surgery in the first year of life.

The literature distinguishes between 2 hypotheses as

the potential explanations for the inverse volume-outcome relationship.⁹ The first hypothesis, practice makes perfect, is based on the supposition that a greater volume of patients should allow the operative team to develop greater skills and judgment in the management of complex clinical conditions and therefore achieve better outcomes. The alternative hypothesis, selective referral patterns, is that operating teams with superior outcomes attract a greater number of patients. These hypotheses have very different policy implications. Under the first hypothesis, and in an environment where cost containment is important, centralization in the least costly institution would be a reasonable strategy; institutions would logically improve outcome with further experience and accomplish this at a lower cost. Obviously, if the second hypothesis holds, then this strategy would not be reasonable, and centralization should be in those institutions with better outcomes. In our view, these hypotheses are not mutually exclusive; a more accomplished team will see selective referral and will show a greater improvement in outcomes as a result of the larger practice than would a less accomplished team. Therefore, should future analyses of primary clinical data confirm our findings, we believe that excellence would be served by centralizing infant heart procedures in those institutions with an established track record.

In conclusion, historically the existence of certificate-of-need regulation was probably an important incentive for the profession and the state to create a continuous quality improvement program. This observation raises the question whether certificate-of-need regulation should be extended to other sophisticated surgical programs, such as AAA repair, or whether market forces provide enough of an incentive to develop such quality improvement programs. The policy issues raised by choosing between regulation and market forces are complicated and beyond the scope of this article. Whatever the mechanism, however, the data emerging from this study provide a powerful challenge to the profession to develop programs to monitor quality. Administrative databases, which are easily accessible, can be used to focus efforts. These studies need to be supplemented by careful observation studies with greater clinical detail to form a foundation for quality improvement programs. New York State's quality improvement program in CABG is a major step in the right direction, but there is still room for improvement. First, there is insufficient detail about the processes of care to pinpoint the problem. Second, the focus on short-term mortality should be extended to long-term death outcomes and the symptoms of patients, as

reflected by their functional status and quality of life, particularly in light of the growing emphasis on the appropriate use of surgical procedures. Finally, we need to understand better how the information is used by surgeons, hospitals, the public, purchasers, payers, and referring physicians.

REFERENCES

1. Luff HS, Bunker JP, Enthoven AC. Should operations be regionalized? An empirical study of the relation between surgical volume and mortality. *N Engl J Med* 1979;301:1364-9.
2. Hannan EL, Sui AL, Kumar D, Kilburn H, Chassin MR. The decline in coronary artery bypass graft surgery mortality in New York State: the role of surgeon volume. *JAMA* 1995;273:209-13.
3. Hannan EL, O'Donnell JF, Kilburn H, Bernard HR, Yazici A. Investigation of the relationship between volume and mortality for surgical procedures performed in New York State hospitals. *JAMA* 1989;262:503-10.
4. Hannan EL, Kilburn H, Bernard H, O'Donnell JF, Lukacik G, Shields EP. Coronary artery bypass surgery: the relationship between in-hospital mortality rate and surgical volume after controlling for clinical risk factors. *Med Care* 1991;29:1094-107.
5. Hannan EL, Kilburn H, O'Donnell J, Lukacik G, Shields EP. Adult open heart surgery in New York State. *JAMA* 1990;264:2768-74.
6. Hannan EL, Kilburn H, Racz M, Shields E, Chassin MR. Improving the outcomes of coronary artery bypass surgery in New York State. *JAMA* 1994;271:761-6.
7. Jenkins KJ, Newburger JW, Lock JE, Davis RB, Coffman GA, Iezzoni LI. In-hospital mortality for surgical repair of congenital heart defects: preliminary observations of variation by hospital caseload. *Pediatrics* 1995;95:323-30.
8. Hannan EL, Racz M, Kavey RE, Quaegebeur JM, Williams R. Pediatric cardiac surgery: the effect of hospital and surgeon volume on in-hospital mortality. *Pediatrics* 1998;101:963-9.
9. Hannan EL, Kilburn H, O'Donnell FJ, Bernard HR, Shields EP, Lindsey ML, Yazici A. A longitudinal analysis of the relationship between in-hospital mortality in New York State and the volume of abdominal aortic aneurysm surgeries performed. *Health Services Research* 1992; 27:517-42.
10. Lemeshow S, Hosmer D. A review of goodness-of-fit statistics for use in the development of logistic regression models. *Am J Epidemiol* 1982;115:92-110.
11. Sollano JA, Sydorak RM, Heitjan DF, Gelijns AC, Reemtsma K, Moskowitz AJ. Relating volume to outcome in a risk-adjusted prediction model for AAA surgery in New York State. *Circulation* 1997;96(Suppl):I1-12.

Discussion

Dr Jack M. Matloff (*Los Angeles, Calif*). This topic has been a recurring one. New York State has frequently been the venue for exploring the thesis that volume and clinical outcomes of a variety of procedural therapies can be related. It is therefore not surprising that one can find support in the literature for almost any relationship. For cardiothoracic surgeons there have been personal biases about the assumptions inherent in SPARCS in the past and the modeling that influences the conclusions drawn. Quite apart from that, your observa-

tion that the data are administrative or observational underlies one of the real problems and, that is, that the results are subject to confounding as an expression of variability of case mixes. This is particularly important because other studies have shown that the more case mixes are adjusted, the smaller are the apparent benefits of increased volumes of operations. I also find myself wondering whether or not the juxtaposition of the 3 groups of patients and the differences in numbers created a problem for you at any time and whether or not you ever considered just studying 1 of the cohorts.

Now, having made these observations, I would like to turn to the reasons that I think we have to be very careful about such studies. Justification for concern can be appreciated from two juxtaposed quotes, one from Victor Fuchs at Stanford, noted health care economist, and the other from Oscar Wilde. Dr Fuchs notes that "the only independent variable in the equation for calculating the cost of health care is the volume of services given." So, volume clearly is an important issue on the price side. By the same token, Oscar Wilde has made the companion observation that we should "avoid the person who knows the cost of everything and the value of nothing."

We do not now have and probably never will have enough resources to provide all the needed health care, so rationing is an invariable consequence. The real issue is the "how" of resource allocation. Controlling volumes through market approaches as espoused by managed care advocates can control costs, but because health care does not proceed in a free market, that is, it is increasingly regulated, one has to be concerned about whether or not current market strategies will be as successful as managed care proponents wish in regard to controlled volumes, cost containment, and quality.

What has to be guarded against is the attempt to justify limiting services on the basis that decreased volume inevitably equals decreased quality. This is a quantum leap that some health care policy makers—and I do not mean to imply in any way that you have participated in this—have made in the past. It is their concern that really is at issue here because this has been done without clinical experience to appropriately support their positions.

Your paper concludes with a modest proposal for creating quality improvement through programs based on clinical data for complex surgical procedures, and it is your hope that the profession would take the initiative in organizing such programs. I repeat this because I think that really is a very modest request and certainly has to be taken in the context of what President Loop spoke about yesterday.

I have a question in regard to whether or not your team has considered a strategy for accomplishing these modest goals. If you were to have such a strategy, would you use numerical quotas?

Finally, last night, while speaking with Dr Favaloro, I was struck by the realization of how far we have come in cardiac care, driven by cardiac surgeons over the past 30 years. This study documents that a CABG operation in many if not most hospitals where it is practiced can be the safest surgical therapy given there. This takes into account the descriptors of our patient populations. If we, as a specialty, have helped accom-

plish this, then we should be capable of regaining the health policy mantle that has slipped from our hands over the past 25 years. I fully expect within the next 5 to 7 years that Dr Reemtsma and his colleagues will return to this forum and will give us the definitive resolution of the issue of volumes and outcomes.

Dr Reemtsma. We have not addressed rationing of care. However, I would like to comment on your point about the strategy for accomplishing these goals. I do not pretend that the data we have submitted are definitive. I suggest that this is an indicator of the kind of information we need to make these decisions. Then the strategy for accomplishing this must come through some kind of professional and public effort. I would cite the Cardiac Advisory Committee in New York as an example. I am skeptical about the role of government alone, uninformed by professional information, to make these decisions, because of the power of economic and political forces that often dictate such policy.

Dr Frank C. Spencer (*New York, NY*). As Dr Reemtsma has already mentioned, I want to acknowledge the leadership of Dr George Humphreys for starting the first Cardiac Advisory Committee in New York State in the 1950s. Dr Humphreys was chairman of surgery at Columbia University at that time. After the enactment of the Crippled Children's legislation, Dr Humphreys was asked to form a committee to assess the competence of different hospital centers for performing congenital heart operation, looking at their staff and their facilities.

He asked me to assume his responsibilities as chairman of the committee in the late 1960s. I subsequently spent about 15 years with the committee and the subsequent State Cardiac Advisory Committee.

I want to emphasize strongly Dr Reemtsma's comment about the great benefit from this form of collective study. The unique value of the New York model is the fusion of the voluntary surgeons and cardiologists with the State Health Department. The voluntary physician component, performed without reimbursement, provides the clinical judgment to advise the state about what action, if any, is indicated. The cardiac committee is purely advisory to the state. This combination avoids the problem of creating a monopoly in violation of the Sherman Antitrust Act. The vast majority of the recommendations given by the committee are subsequently followed by the State Health Department.

A major achievement in the past several years has been developing the risk-adjusted mortality rates for coronary bypass. The benefit for everyone, as emphasized by Dr Reemtsma, is that the collective results define what is achievable. This collegial sharing of information is invaluable. For example, if a center had a mortality rate of 5% with a certain type of bypass, although comparable centers achieved a mortality rate of 2% to 3%, there is an automatic stimulus to improve.

The hazard that evolved in New York State is the inability to keep the data confidential. This was confidential for over 20 years, but a law suit by newspapers based on the Freedom of Information Act resulted in the state data becoming accessible to the public. The Society of Thoracic Surgery on a

national level has subsequently developed a much broader database that should be emulated. To repeat, the goal is to define what is achievable.

What volume of annual operations was necessary for the mortality rate with CABG to plateau? The data presented show a mortality plateau, but the centers are required to have a minimal number of cases per year. I believe the minimum is at least 200 cases per year. The question nationally is what minimal number of cardiac operations would a center perform each year to achieve good results.

In the future, similar information can be sought for outer complex operations, such as thoracic and abdominal aneurysm.

Dr Reemtsma. I agree that your experience with the Cardiac Advisory Committee is one that gives us encouragement for this blend of voluntary and regulatory control.

In regard to your question about what level of experience is required to reach that plateau, this is not answered definitively by our data. In the CABG operation it is probably somewhere between 200 and 400 patients per year. In aneurysms, it is probably somewhere in the range of 20 to 40 cases per year.

Dr Craig R. Smith (*New York, NY*). Although this study did not identify a volume threshold for good results in CABG operations in a regulated, high-volume environment like New York, that does not mean that such a threshold does not exist in an unregulated, high-volume environment. Los Angeles comes to mind, with more than 50 heart teams serving a population less than one half that of the state of New York, as a place where volume thresholds might be detectable. Have you thought about addressing this question in other regions, where coronary surgery is unregulated?

Dr Reemtsma. I think that is an important question and one that we will address. You are correct that New York represents the extreme in regulation of cardiac surgical procedures and has for years functioned under the certificate-of-need process. California, by contrast, is an example of a laissez-faire system. We are looking at these differences, but we do not have the data on this yet.

Dr Charles Marrin (*Lebanon, NH*). The Northern New

England Cardiovascular Disease Study Group has wrestled with the issues described in Dr Reemtsma's paper for more than 10 years. It is really wonderful to hear his presentation.

It is easy to pick holes in studies of this kind, and we all have at one time or another. But the truths contained in this article cannot be denied. I think it is important to remember that he is not one of *them*, he is one of *us*, and he has led the way. It is time for the rest of us to get beyond anger and denial and to embrace outcome studies. Rather than quibbling about methods of studies done by *them*, let *us* do the studies ourselves and do them right. That way we will believe the data.

Although outcome studies are not going to go away, the question is will they be done to us or by us? We can control our destiny, and the choice is ours.

Dr Reemtsma. It is a pleasure to pay tribute to the Dartmouth group for the outstanding work that you have done, and I appreciate your comments.

Dr Jacob Kolff (*Johnstown, Pa*). In Pennsylvania, the hospital cost-containment council has reported hospital mortality rates for surgeons and for hospitals for CABGs in 1991, 1992, and 1993. In 1993, you can divide up the hospitals into 3 groups: 13 centers doing 60 to 289 procedures with a mortality rate of 3.8%; the middle group from 290 to 600 procedures with a mortality rate of 2.9%; and the larger centers, 9 centers, doing 600 to 954 CABGs per year, with a low mortality rate of 2.5%.

Can we qualitatively afford to have gas stations on every corner now that the certificate-of-need requirements in Pennsylvania, and I think also in Ohio, have been eliminated?

Dr Reemtsma. The answer to your question is that the certificate-of-need process has survived in New York, and I think it has served our profession and patients well. I believe that what you are describing, the relationship between mortality rates and outcome, was more pertinent in New York before 1990.

We have seen data on every hospital in the state of New York that does cardiac operations, including volume, mortality rates, and risk-adjusted mortality rate. This information was collected under auspices of the Cardiac Advisory Committee.