The Volume-Outcome Relationship: Practice-Makes-Perfect or Selective-Referral Patterns?

Harold S. Luft, Sandra S. Hunt, and Susan C. Maerki

Various studies have demonstrated that hospitals with larger numbers of patients with a specific diagnosis or procedure have lower mortality rates. In some instances, these results have been interpreted to mean that physicians and hospital personnel with more of these patients develop greater skills and that this results in better outcomes—the "practice-makes-perfect" hypothesis. An alternative explanation is that physicians and hospitals with better outcomes attract more patients—the "selective-referral pattern" hypothesis. Using data for 17 categories of patients from a sample of over 900 hospitals, we examine the patterns of selected variables with respect to hospital volume. To explore the plausibility of each hypothesis, a simultaneous-equation model is also used to test the relative importance of the two explanations for each diagnosis or procedure. The results suggest that both explanations are valid, and that the relative importance of the practice or referral explanation varies by diagnosis or procedure, in ways consistent with clinical aspects of the various patient categories.

This research was supported by Grant Numbers HS-04329 and HS-05411 from the National Center for Health Services Research, Department of Health and Human Services. An earlier version of this article was presented at the American Public Health Association Annual Meeting in Anaheim, CA, November 1984.

Address correspondence and requests for reprints to Harold S. Luft, Ph.D., Professor of Health Economics, Institute for Health Policy Studies, School of Medicine, University of California, San Francisco, 1326 Third Avenue, San Francisco, CA 94143. Sandra S. Hunt, B.A., Research Assistant at the Institute for Health Policy Studies, University of California at San Francisco, at the time this research was done, is now a Health Benefits Consultant with Coopers & Lybrand in San Francisco. Susan Maerki, M.H.S.A., M.A.E., a Senior Research Associate at IHPS at the time of this research, is now a Corporate Planning Analyst with Blue Cross of California.

A growing number of studies focus on the relationship between the number of patients treated with a specific diagnosis or procedure at a hospital and the clinical outcomes of those patients. Most of these studies concentrate on surgical procedures, such as coronary artery bypass graft, total hip replacement, and prostatectomy [1, 2]. Some studies also include medical diagnoses, such as acute myocardial infarction, peptic ulcer disease, subarachnoid hemorrhage, and burns [3-5]. In many, outcomes are measured by mortality rates during hospitalization, although some authors also consider morbidity, length of stay, and mortality up to 60 days after discharge [6-9].

While it is difficult to combine findings across studies because of differences in data, sample, outcome measures, and methodology, all have found an inverse relationship between high patient volume and poor outcomes for at least some categories of patients. Authors and policymakers have sometimes interpreted these results to mean that higher volumes lead to better results [10]. There is certainly a logical argument in favor of that interpretation. In other industries there is evidence of a learning curve whereby production becomes more efficient with greater experience [11, 12]. Similarly, one would expect a surgical team that performs one open heart procedure a day to be more proficient at it than a team that performs one a month.

On the other hand, some analysts argue that the observed relationship may reflect a referral system that channels more patients to hospitals and physicians that achieve better results [13]. It is plausible that a referral system could function without explicit knowledge of objective outcomes by either patients or primary care physicians. It could be that specialists and hospitals develop a reputation as "the best in the area" for a particular type of case based upon low complication rates or the general perception that patients have better outcomes there. Alternatively, some providers may develop poor reputations among referring physicians and therefore lose referrals.

The two causal models have substantially different policy implications. If the observed patterns reflect only the "practice-makes-perfect" phenomenon, then any effort to concentrate patients in selected hospitals will improve outcomes. Thus, Medicaid programs or preferred provider organizations that obtain low prices from hospitals in return for more patients would be justified on grounds of both economy and quality [14, 15]. On the other hand, if the observed patterns are entirely due to intelligent selective referrals, where patients are channeled to the best providers, then contracting on the basis of price will not result necessarily in beneficial patient care outcomes.

This article will use a patient abstract data set to explore the practice-makes-perfect and selective-referral pattern hypotheses as alternative, but not mutually exclusive explanations of the inverse relation between volumes and outcomes. Given the substantial differences in policy implications of the two models, it is important that our analysis be as robust as possible. Therefore, we will use several different types of evidence, ranging from simple tables to simultaneous-equation models. Furthermore, we are not attempting total rejection of either hypothesis. Our intent is to build a plausible case that, for some diagnoses and procedures, selective-referral patterns need to be considered along with the practice-makes-perfect effect.

ALTERNATIVE EXPLANATIONS OF THE OBSERVED VOLUME-OUTCOME RELATIONSHIP

The practice-makes-perfect explanation of the observed volume-outcome relation rests on the general notion that increased experience results in more finely developed skills and, therefore, in better outcomes. Evidence from various industries shows that production costs per unit fall as a firm accumulates experience [11, 12]. This differs from medical care in that such studies focus on accumulated experience—for example, the cost of building the first, tenth, hundredth, and thousandth fighter-bomber with each unit essentially of the same design. While accumulated experience is related to volume per unit time, such as a year, the relationship is complex. Clearly, a surgeon with a high operative load develops experience much faster than one with a low volume, but cumulative experience can vary widely at given yearly volume levels. Thus, one would expect age to be a crucial variable in analyses with physician-specific data, yet Flood, Scott, and Ewy [10] find no relation between outcome and surgeon characteristics.

The direct analogy between manufacturing's "learning curve" and development of a particular medical care skill is probably most valid when procedures and treatments are still in the experimental phase. In both cases, new approaches are being developed and general skills are being applied to specific circumstances. However, once a technique is perfected, it can often be taught fairly readily. Although each person has an individual learning curve, the flat portion of the curve is probably attained during the training period. Thus, experience accumulated over many years is not a likely explanation of the observed relation.

A more likely aspect of experience is the deterioration of skills with lack of practice. In comparing two equally trained surgical residents, the one who consistently performs many of the procedures in question will maintain—or continue to improve—his or her skills, while the one who performs few such procedures will become progressively less proficient. This model of skill maintenance is intuitively plausible and can account for other factors that may explain the volume-outcome relation. For example, nursing and other staff who are more familiar with certain types of patients may become or remain more proficient in working with them. Higher volumes may make it possible, also, for hospitals to purchase specialized equipment for such patients [9]. This explanation is closer to "economies of scale" than to "practice makes perfect," but the two are similar in this context.

Unfortunately, the practice-makes-perfect model does not explain why some hospitals and physicians have high volumes of certain types of patients and others have low volumes. Obviously, the size of the hospital is likely to be a predictor of diagnosis- or procedure-specific volumes, primarily because hospital size will be a good indicator of the market area from which patients are drawn. Other things equal, a 400-bed hospital will serve a population about twice as large as a 200-bed hospital, and the number of people in its area requiring any procedure, such as prostatectomy, will be twice as large. Hospitals with large medical staffs relative to their size may also have higher volumes because of a larger potential pool of patients. Once one controls for hospital size, the larger medical staff will have a greater number of patients. This assumes an average patient load per physician, although not necessarily at that hospital if the physicians have multiple hospital affiliations or staff privileges. In these circumstances, the pool of potential patients with specific diagnoses will increase. Hospitals with major teaching affiliations may have above-average volumes of patients requiring specialized services or facilities. Hospitals may develop programs or draw subpopulations at high risk for certain conditions, such as cirrhosis in inner cities and leg fractures in ski areas. In addition to such straightforward predictors of volume, one must ask whether hospitals or physicians with better results might attract more patients than would otherwise be the case—the selective-referral pattern hypothesis.

The notion that patients in some instances may look for the hospital or physician with the best results seems implausible to some. Flood, Scott, and Ewy reject the selective-referral hypothesis largely because they think the variation in mortality by disease is too small to influence patient choice [10]. However, if complications are correlated with mortality (which is often the case), variations in outcomes may be large enough to be noticeable—if not by patients, then by their primary physicians who choose specialists for referral. While it is difficult to identify an individual hospital or physician as having significantly worse-than-average death rates [16], referral patterns may be based on a far more simple set of decision rules. If primary physicians initially choose specialists at random, then switch referrals after one "bad outcome," patients eventually are directed away from providers with outcomes truly worse than average. Furthermore, even if the majority of patients go to the nearest hospital or otherwise make decisions independent of perceived outcomes, a minority seeking or referred to the "best provider in town" (or referred away from "poor-quality providers") will result in a selective referral pattern for specific diagnoses and procedures. This would mean that hospitals with better outcomes will have higher-than-expected volumes. The question, therefore, is whether *some* patients are influenced in their choice of physicians and hospitals by relative performance, not whether all patients are so influenced.

A principal empirical objection to the selective-referral hypothesis is that some studies show little relationship between outcomes and those variables traditionally considered to be markers of good performance, such as teaching status of hospitals or board certification of physicians [10, 17]. While it is possible that such characteristics may also be associated with unmeasured increases in risk factors, these measures are a rather blunt indicator of special expertise. An alternative explanation is that there is substantial variability in performance by providers *within* groupings. This is more reasonable, because we are attempting to measure diagnosis- or procedure-specific performance. It is not uncommon for a teaching hospital to be outstanding in the treatment of a selected diagnosis or procedure—for example, cardiovascular surgery—but not to be particularly distinguished in another, such as neurosurgery.

If externally measured variables are insensitive to differences in outcomes which influence referrals, then higher-than-expected volumes for a specific procedure or diagnosis may, in fact, be the best single indicator of exceptionally good outcomes. If so, at least part of the causal linkage is from outcomes to volumes. As an analogy, consider the situation of a new visitor to a city. The hotel's restaurant guide provides a description of the local options, including the type of food, price range, and where the chef trained. This may be of some guidance in finding the best restaurant, but our visitor is likely to continue to be uncertain. A better indication of relative quality might be the number of patrons in each establishment. Our visitor would probably be wise to avoid places that are nearly empty and, if there is no hurry, a long line would be not only a measure of popularity but perhaps the best single indicator of good food within a given price range. One would not, however, argue that the food is good simply because the lines are long. Of course, the restaurant example raises many other issues—one is more likely to be impressed by an establishment filled with locals than one full of out-of-towners, or by a busy place in an out-of-the-way location.

To further refine the distinctions between practice-makes-perfect and selective-referral patterns, the former suggests that, while high volumes are associated with better outcomes, there is no reason to expect that volumes are related to patient risk factors or the proportion of patients transferred into the hospital. The selective-referral hypothesis, on the other hand, directly implies that high-volume settings should attract transfers from other hospitals and have a higher number of patients with the procedure or diagnosis than would be expected based on variables such as hospital size. Furthermore, we would expect selective referrals to be more important for nonemergency, but risky conditions, especially those generally treated by referral specialists rather than by primary care physicians.

With data from a broad sample of U.S. hospitals, the validity of the explanations for the observed volume-outcome relationships can be tested by either of two empirical approaches. One is to explore whether certain characteristics, such as the proportion of patients transferred in or out of a hospital, vary with volumes in patterns consistent with either hypothesis. A second relies upon simultaneous-equation techniques to test the relative importance of selective-referral and practice-makes-perfect effects while holding patient and hospital characteristics constant. The basic data used are described in the next section, and the subsequent two sections present the results of each approach.

Data and Case-Mix Adjustment

The data for this study are drawn from several sources. Patient volume and outcome data were derived from case abstracts, as described in the next subsection. These data were also used to develop measures of case-mix-adjusted expected mortality rates for each hospital. Additional data concerning the hospital and its environment were drawn from other sources.

PATIENT SELECTION AND CASE-MIX ADJUSTMENT

The patient data are based on discharge abstracts from the Professional Activities Study of the Commission on Professional and Hospital Activities (CPHA) in Ann Arbor, Michigan. All United States hospitals subscribing to the system for the entire year of 1972 were eligible. To be included, a hospital had to admit at least one patient with any of the 17 selected diagnoses or procedures. Patient records were excluded if data were missing on age, discharge status, or length of stay. After this patient selection, all information was aggregated by hospital, and analysis proceeded on a diagnosis- or surgical procedure-specific basis.

Because of confidentiality restrictions, we were not permitted access to individual patient records. To adjust for case mix, CPHA classified patients according to characteristics known to be associated with differential mortality rates; these included age, sex, single or multiple diagnosis, admission blood pressure, and, for respiratory distress syndrome, birth weight. The number of classification cells ranged from 6 for appendectomy, which was divided by age group and single/multiple diagnosis, to 18 for abdominal aortic aneurysm and acute myocardial infarction, which were split by age group/admission blood pressure/sex, and age group/admission blood pressure/single or multiple diagnosis, respectively (a $3 \times 3 \times 2$ matrix). (Specific breakdowns for the cells are available in an appendix from the authors.) Cells were examined to make sure that cell-specific death rates were not based on denominators too small to be reliable. In only a few instances were cells combined due to a small number of observations.

ACTUAL AND EXPECTED MORTALITY RATES

National and individual hospital mortality rates for each classification cell within a diagnosis or procedure are simply the number of in-hospital deaths among patients in the cell divided by the number of patients in the cell. These United States averages were based on all patients in each of our diagnosis or procedure groups to increase the reliability of the figures. Later we dropped data for 356 hospitals that did not match a national hospital file providing additional independent variables. The final sample of hospitals is all U.S. community hospitals, except for an overrepresentation of larger hospitals and those in the North Central states. The resulting biases are not likely to affect the questions addressed in this article, because our focus is on the volume-outcome relations within a group of hospitals, not on the magnitude of the volume-outcome impact for the United States as a whole. Figures for the number of hospitals, patients, and deaths for each group are listed in Table 1. The final hospital-based mortality statistics for the 17 diagnoses and procedures reflect the experience of 1,008,502 patients admitted to hospitals in 1972.

Table 1: Hospital and Patient Statistics—1972

	Number of Hospitals	*Number of Patients*	*Number of Deaths*	*HICD-8 Code Numbers*
Abdominal aortic aneurysm (with or without rupture)	736	6,065	1,530	441.3-441.4
Acute myocardial infarction	906	98,066	15,925	410.0-410.5, 410.8-410.9
Cirrhosis	913	24,228	3,412	571.0-571.2, 571.8-571.9
Fracture of the femur	910	46,468	3,675	820.0-820.5
Peptic ulcer	913	142,870	1,503	531.0-531.1, 532.0-532.1, 533.0-533.1, 535.0
Respiratory distress syndrome	770	16,373	3,241	776.1-776.2, 777.0-777.1
Subarachnoid hemorrhage	749	5,049	2,210	430.0-430.1
Angiography and cardiac catheterization	360	26,678	479	32.0, 95.2
Appendectomy	916	80,211	273	49.1
Cardiac bypass	114	5,172	312	30.5
Cholecystectomy	914	102,917	806	53.5
Inguinal hernial repair	920	134,497	274	57.0, 57.1
Hysterectomy (abdominal and vaginal)	915	180,464	267	71.0-71.4
Intestinal operations (colectomy, ileostomy, colostomy, anastomosis, protectomy)	898	36,860	3,258	47.0-47.5, 48.1-48.2, 50.2
Stomach operations (gastrectomy, vagotomy, colectomy)	864	24,072	1,344	44.3-44.6, 45.3, 45.5
Total hip replacement	730	20,429	1,070	84.1
Transurethral prostatectomy	756	58,083	657	65.2

The actual hospital mortality rate is the total number of in-hospital deaths divided by the number of patients admitted for that diagnosis or procedure. To adjust for case mix, an indirectly standardized expected death rate was computed. This expected death rate is the sum over all classification cells of the national cell-specific death rate for a given diagnosis or procedure multiplied by the proportion of patients in the hospital who fell into the relevant cell. The expected death rate for a hospital is the death rate that would result if the hospital's death rate for each patient category was equivalent to the national average death rate for that cell, weighted by its own mix of patients. The outcome experience of each hospital is measured by comparing the actual death rate (ADR) to the expected death rate (EDR).

OTHER SOURCES OF DATA

Additional information on the patients, specialized equipment, medical staff, sociodemographic factors, and number of neighboring hospitals was used in the analysis. The CPHA patient abstracts provided such diagnosis- or procedure-specific measures as transfers in and out of the hospital, percent with surgery within six hours of admission, units of blood transfused, percent with diagnostic tests such as gastroscopy, or percent with abdominal rather than vaginal hysterectomy. Information on the hospital and medical staff and on specialized equipment that would indicate modern, perhaps better-quality, care were taken from the 1972 Survey of Specialized Clinical Services and the 1973 Survey of Medical Staff Organization. This information included number of beds in special care units, medical school affiliation, and the ratio of staff physicians to beds. Area sociodemographic variables are from the 1972 Area Resource File, and the count of neighboring hospitals was generated for other portions of our research [18].

For each diagnosis or procedure, histograms were generated showing the number of hospitals with specific levels of patients. Cutoffs were chosen for very low, low, moderate, high, and very high volumes based on two criteria: (1) the pattern of hospitals across small increments in volumes and (2) a desire to have a reasonable proportion (10–30 percent) of the hospitals in each category. Table 2 presents the volume cutoffs for each volume level.

Table 2: Volume Cutoffs for Each Volume Grouping

	Very Low	*Low*	*Moderate*	*High*	*Very High*
Abdominal aortic aneurysm	1-2	3-6	7-12	13-24	25 +
Acute myocardial infarction	1-12	13-24	25-120	121-240	241 +
Cirrhosis	1-4	5-18	19-36	37-72	73 +
Fracture of the femur	1-12	13-36	37-72	73-120	121 +
Peptic ulcer disease	1-24	25-72	73-180	181-240	241 +
Respiratory distress syndrome	1-5	6-15	16-35	36-72	73 +
Subarachnoid hemorrhage	1	2-3	4-6	7-12	13 +
Cardiac catheterization	1-4	5-50	51-150	151-400	401 +
Appendectomy	1-12	13-24	25-72	73-180	181 +
Cardiac bypass graft	1-4	5-24	25-72	73-200	201 +
Cholecystectomy	1-12	13-24	25-72	73-180	181 +
Hernia repair	1-12	13-24	25-72	73-180	181 +
Hysterectomy	1-24	25-120	121-240	241-360	361 +
Intestinal operations	1-6	7-24	25-48	49-96	97 +
Stomach operations	1-6	7-18	19-36	37-60	61 +
Total hip replacement	1-2	3-24	25-48	49-72	73 +
Transurethral prostatectomy	1-6	7-24	25-120	121-180	181 +

DIFFERENTIAL PATTERNS WITH RESPECT TO VOLUME

One simple approach to exploring the relation between volume and outcome is to categorize hospitals by the number of patients in a particular diagnosis or procedure category and then to examine the patterns for selected variables across volumes and types of patients. Table 3 presents the simple relation between volumes and outcomes, as measured by the ratio of actual to expected death rates for hospitals in various volume categories. For all but three of the procedures or diagnoses, outcomes tend to improve (actual divided by expected death rates fall) in hospitals with higher volumes. The figures represent total deaths to total expected deaths for all patients in hospitals within the

Table 3: Actual/Expected Mortality, by Volume in Hospital

	Very Low	*Low*	*Moderate*	*High*	*Very High*
Decreasing Pattern					
Cardiac bypass graft	1.466	2.051	1.317	0.881	0.629
Total hip replacement	1.441	1.382	1.180	0.702	0.773
Transurethral prostatectomy	2.462	1.385	0.983	0.963	0.957
Intestinal operations	1.407	1.233	1.033	0.941	0.868
Stomach operations	1.317	1.158	1.069	0.932	0.876
Hysterectomy	1.874	1.471	1.125	0.895	0.733
Cholecystectomy	1.250	1.375	1.375	1.110	0.745
Abdominal aortic aneurysm	1.147	1.258	1.106	0.968	0.730
Appendectomy	1.233	1.823	1.070	0.980	0.826
Acute myocardial infarction	1.274	1.121	1.078	0.985	0.923
Cirrhosis	1.118	1.147	1.014	0.982	0.978
L- or U-Shaped Pattern					
Cardiac catheterization	4.214	1.826	1.000	0.813	1.231
Hernia repair	1.458	1.586	1.028	1.000	1.000
Respiratory distress syndrome	1.554	1.282	0.898	0.925	0.945
No Clear Pattern					
Peptic ulcer	0.178	1.000	1.097	1.040	0.957
Fracture of femur	0.716	1.108	0.972	1.000	0.930
Increasing					
Subarachnoid hemorrhage	0.802	0.899	0.921	1.037	1.091

volume category—that is, the experience of all patients in hospitals in a given category, not the average of the hospital-specific ratios. Although estimates for the "very low" category seem sensitive to small sample problems, the observable patterns provide fairly convincing evidence of a relationship between volume and outcome. Given the simiiarity between these findings and the others in the literature, we can proceed to explore the causal relationships.

TRANSFERS INTO A HOSPITAL

A direct measure of selective referrals is the proportion of patients transferred into a hospital from another acute hospital. While the groupings of diagnoses and procedures in Table 4 may be subjective, it is clear that there are marked differences in transfer rates with respect to volume. The average transfer rate is less than 2 percent of the patients with respiratory distress syndrome (RDS) or subarachnoid hemorrhage in very low volume hospitals but 11–17 percent of those in very high volume hospitals. (For these two diagnoses the very high volume hospitals average more than twice as many patients as the next lower group. Even if all the transfers were deleted, they would still be very high volume settings.)

Six of the procedures are roughly flat and show little relation between volume levels and transfer-in rates. Hysterectomy, peptic ulcer, and acute myocardial infarction (AMI) apparently exhibit a curious L-shaped pattern, with high rates of transfers in very low volume hospitals, followed by no pattern thereafter. These may be artifacts of the data; for hysterectomy, the "twice normal" rate reflects one patient transferred. Inspection of the individual hospital data for ulcer suggests a data problem, with fractional transfers reported for some hospitals and 16 of 24 patients transferred into another. The "high" transfer rates for AMI are due to a number of hospitals with two of six, two of three, or three of four patients being transferred in. In several instances, these are small rural hospitals, and it is likely that a post-infarction patient was "returned home" from another setting.

TRANSFERS TO OTHER HOSPITALS

The flip side of the transfer-in rate is the proportion of patients discharged to another hospital, rather than to a convalescent facility or home. As might be expected, for most diagnoses and procedures the transfer-out rate falls with volume (see Table 5). For some types of patients the pattern is quite dramatic, with one-third to almost one-half of the subarachnoid hemorrhage and respiratory distress syndrome patients in very low volume hospitals being transferred, in contrast to 5–8 percent for very high volume hospitals. The very high rate of transfers for subarachnoid hemorrhage and RDS suggests emergency admissions to the nearest hospital or place of birth for RDS, followed by subsequent transfer to a more appropriate setting. With the exception of cardiac catheterization, these groups are all medical rather than surgical admissions. Some of the patients with cardiac catheterization

Table 4: Proportion of Patients Transferred In by Volume in Hospital

	Very Low	*Low*	*Moderate*	*High*	*Very High*
Strong Increasing Pattern					
Respiratory distress syndrome	.015	.014	.014	.064	.171
Subarachnoid hemorrhage	.014	.027	.031	.072	.114
Abdominal aortic aneurysm	.006	.005	.019	.029	.037
Weak Increasing Pattern					
Stomach operations	.003	.006	.005	.009	.009
Intestinal operations	.002	.007	.010	.007	.013
U-Shaped					
Transurethral prostatectomy	.027	.003	.006	.007	.012
Cirrhosis	.019	.005	.007	.010	.011
Fracture of femur	.042	.021	.028	.033	.041
"Roughly Flat"					
Total hip replacement	.017	.016	.019	.024	.022
Cardiac bypass graft	.013	.020	.041	.019	.025
Cardiac catherization	.016	.016	.013	.012	.012
Hernia repair	.000	.000	.001	.001	.001
Cholecystectomy	.002	.002	.002	.002	.002
Appendectomy	.000	.001	.003	.002	.001
L-Shaped					
Hysterectomy	.0008	.0004	.0002	.0004	.0004
Peptic ulcer	.077	.003	.003	.004	.003
Acute myocardial infarction	.060	.009	.009	.007	.009

in very low volume hospitals may not have been admitted for that purpose.

EXPECTED DEATH RATES

A final piece of evidence with respect to selective referrals is the risk pattern of patients. An expected mortality rate based on patient char-

Table 5: Proportion of Patients Transferred Out by Volume in Hospital

	Very Low	Low	Moderate	High	Very High
Strong Decreasing Pattern —High Rates					
Subarachnoid hemorrhage	.464	.284	.215	.123	.051
Respiratory distress syndrome	.337	.220	.197	.143	.078
Abdominal aortic aneurysm	.197	.075	.032	.015	.014
Fracture of femur	.239	.073	.037	.033	.042
Strong Decreasing Pattern —Low Rates					
Acute myocardial infarction	.063	.051	.026	.010	.009
Peptic ulcer	.037	.012	.006	.004	.003
Cardiac catheterization	.043	.034	.012	.012	.006
L-Shaped or Weak					
Intestinal operations	.023	.010	.012	.007	.006
Stomach operations	.009	.010	.007	.006	.004
Hysterectomy	.0039	.0016	.0007	.0004	.0004
Hernia repair	.007	.002	.002	.001	.001
Cardiac bypass graft	.026	.010	.008	.011	.0003
Cholecystectomy	.008	.004	.004	.002	.001
Cirrhosis	.073	.044	.022	.015	.022
No Clear Pattern					
Transurethral prostatectomy	.009	.004	.004	.004	.004
Appendectomy	.000	.002	.003	.002	.001
Total hip replacement	.022	.026	.022	.018	.019

acteristics can be computed for each hospital. As seen in Table 6, for about half the patient categories there is no pattern of increasing risk with respect to volume. Hospitals with a high volume of RDS tend to have higher-risk patients, as might be expected for a diagnosis subject to extensive regionalization efforts. Perhaps surprisingly, some diagnoses and procedures exhibit a decline in patient risk with respect to volume. One explanation of this is that physicians in high-volume

Table 6: Expected Death Rate by Volume In Hospital

	Very Low	*Low*	*Moderate*	*High*	*Very High*
Strong Increasing Pattern					
Respiratory distress syndrome	.166	.179	.186	.201	.208
Essentially Flat					
Hysterectomy	.0012	.0016	.0016	.0015	.001
Cholecystectomy	.008	.008	.008	.008	.008
Fracture of femur	.082	.081	.081	.080	.080
Intestinal operations	.089	.090	.091	.091	.091
Transurethral prostatectomy	.012	.014	.012	.011	.011
Cardiac bypass graft	.073	.073	.060	.059	.062
Subarachnoid hemorrhage	.419	.434	.431	.429	.427
Stomach operations	.061	.056	.058	.057	.054
Hernia repair	.003	.005	.002	.002	.002
Decreasing Pattern					
Total hip replacement	.057	.054	.050	.046	.043
Cirrhosis	.149	.142	.137	.139	.136
Abdominal aortic aneurysm	.288	.265	.262	.247	.234
Cardiac catheterization	.028	.020	.017	.016	.019
Peptic ulcer	.012	.012	.011	.010	.009
Appendectomy	.005	.004	.004	.003	.003
Acute myocardial infarction	.207	.215	.202	.194	.194

settings avoid certain procedures, such as total hip replacement or cardiac catheterization, in very high risk patients. A second explanation is that many of the patients with medical problems such as cirrhosis, aneurysm, appendicitis, and acute myocardial infarction seen in very low volume hospitals are emergency admissions. If there had been an opportunity to send the patient elsewhere, perhaps because the patient's risk factors were lower, the patient would have been sent there.

While the various measures sometimes present different results, several important lessons can be learned from these tables. The patterns of transfers-in for some patient categories reflect marked trends consistent with selective referrals and inconsistent with the position that the practice-makes-perfect hypothesis is the only explanation of the volume-outcome relation. Furthermore, the patterns of transfers-out and expected mortality rates suggest that for some patient categories, very low volume hospitals probably receive occasional patients because of emergencies, a result consistent with the selective-referral hypothesis. It also suggests that the notion of selective referrals includes both sending patients *toward* some physicians and hospitals with better results and *away* from others with worse-than-average results.

A SIMULTANEOUS TEST OF THE ALTERNATIVE HYPOTHESES

Because our primary objective is to demonstrate that both practice-makes-perfect and selective-referral patterns are viable—and perhaps simultaneous—explanations of the volume-outcome relation, it is appropriate to use an approach that tests both explanations at the same time. Our simultaneous-equation model uses one equation that explains outcomes as a function of volume and other factors, and a second equation that explains volumes as a function of outcomes and other factors.

Simultaneous equations have long been used in econometric modeling, but non-economists are suspicious of some sleight-of-hand. Statistically, what is necessary is a subset within the set of exogenous variables that can be convincingly excluded from the first equation and another subset that can be excluded from the second equation. If we consider the underlying relationships, it is possible to specify the variables to include and those to exclude.

A TWO-EQUATION MODEL

The practice-makes-perfect hypothesis predicts that increased experience will result in more finely developed skills and thus better outcomes. Yet the argument can be based either on cumulative experience (which no one has yet measured) or on volume per unit of time, usually a year. In consideration of the latter notion, it is probably the long-run volume of a hospital or team that matters, not minor variations due to random occurrences. Thus, if one can specify an equation that explains most of the interhospital variation in volumes as a function of hospital size, specialized facilities, and other factors, then the results of such an equation represent "long-run" volume levels and should be closely associated with observed outcomes. The residuals, or the difference between observed and predicted volumes, on the other hand, should have little impact on outcomes if long-run volumes are what counts in "practice makes perfect." Thus, an equation explaining outcomes should really use long-run or predicted volumes as an independent variable together with the hospital and patient characteristics expected to influence outcomes.

If we switch perspectives and allow the residuals to be interpreted as other-than-random variation around long-run volumes, one obvious interpretation is selective referrals. Hospitals with more patients than expected, given their size, teaching status, and other factors, have more patients because of selective referrals, possibly because they are known for good outcomes. In contrast, hospitals may have lower-than-expected volumes because they are avoided due to poor outcomes. Thus, an equation predicting volumes should have outcomes as one of the independent variables, and this outcome variable should "pick up" some of the residual variation.

Our specification of the exogenous variables includes some variables that are the same for all diagnoses and procedures, and some that are linked to specific types of patients and treatments. Table 7 lists the variables included in the two basic equations. Patient volume is represented in log form. As can be seen from the results in Table 3, there is often a curvilinear relationship between outcome and volume which can be approximated by the semi-log specification in the outcome equation, as discussed further on. The simpler quadratic formulation is not feasible in a simultaneous-equation model. Volume is hypothesized to be a function of the size of the hospital—as measured by a series of

Table 7: Variables Included in Two Basic Equations

Volume	*Outcomes*
Death rate (ADR-EDR)	Log volume
Neighbors 0–15	Northeast
Medical school affiliation	North Central
Medical staff/bed	West
Beds 1–49	
50–99	Medical school affiliation
200–299	Transfers in
300–399	Transfers out
400 +	Medical staff on
Volume of appendectomies	governing board

dummy variables representing bed categories—and the volume of appendectomies. The set of variables captures aspects of size and also of service complexity, which is strongly associated with hospital size once the number of routine admissions, such as appendectomies, is held constant [9, 17]. Another potential influence on volumes is the ratio of physicians on the hospital's medical staff to the number of beds. The higher this ratio, the larger the implicit patient base from which the hospital can draw. In general, high ratios reflect multiple affiliations for physicians, which should make interhospital referrals easier. A dummy variable indicating medical school affiliation is also included because the tertiary care capabilities of a teaching center may increase patient referrals, independent of the effects of teaching programs on outcomes [17]. The number of neighboring hospitals within a 15-mile radius is included to capture the possible effect of potential competitors on splitting the market or on providing referrals. Finally, the actual minus expected death rate in the hospital is used to test whether hospitals with better-than-predicted outcomes attract more patients—the selective-referral hypothesis. This formulation of performance is preferred because the ratio of actual to expected deaths gives the same value for all hospitals with zero deaths, even if expected mortality rates vary markedly.

In the equation explaining outcomes (actual minus expected death rate is the dependent variable), several factors can be expected to influence outcomes, but not volumes. Substantial evidence indicates variations in mortality rates by geographic region [6, 17]. The reasons for these regional differences are not well understood, but they have been observed in many settings and are apparent even in postdischarge mortality [6]. Shortell and LoGerfo [4] found that hospitals with medical staff representation on the governing board have better outcomes possibly because of better "quality control." The transfer-out rate is included to make sure hospitals do not "look good" simply by transferring their sickest patients. The proportion of patients transferred in is used to control for "dumping" and other risk factors not captured in the risk matrix. As indicated above, medical school affiliation appears in both equations. Finally, the log of the number of patients with the procedure or diagnosis is included as an endogenous variable.

For each procedure or diagnosis, other variables representing factors expected to attract patients may be included. Examples are presence of a coronary care unit for acute myocardial infarction patients, or variables capturing risk factors, such as percent of patients receiving blood, specific invasive tests, or emergency surgery. Several hospital characteristics such as proprietary and governmental ownership were found to have no effect in either equation and were therefore omitted in the final models. Other variables often included in analyses of hospital utilization were not tested. Insurance coverage and per capita income in the area were not tested because they are likely to influence primarily the population-based hospital use rate. To understand their role in the choice of *which* hospitals attract specific patients, one would require data from all hospitals in a geographic area.

Table 8 presents the estimated coefficients for acute myocardial infarction patients as an example of the type of results one obtains from

Table 8: Acute Myocardial Infarction

Log Volume			*ADR-EDR*		
Constant	2.343	11.39***	Constant	0.105	4.92
Death rate (ADR-EDR)	-0.786	-0.61			
Neighbors 0–15	0.0003	5.99***			
Medical school affiliation	-0.022	-0.51	Medical school affiliation	-0.006	-1.21
Medical staff/ bed	0.0008	0.92			
Beds 1–49	-0.468	-2.48*			
50–99	-0.212	-2.32*			
200–299	0.317	5.83***	Log volume	-0.019	-4.54***
300–399	0.444	7.48***	Northeast	-0.002	-0.37
400 +	0.587	9.43***	North Central	0.003	0.61
Volume appendectomies	0.402	10.99***	West	-0.046	-6.41***
Any CCU/ICU	0.367	2.71**	Transfers in	0.08	-0.57
R^2	0.61		Transfers out	-0.041	-0.42
DF	632		Med staff on governing board	-0.009	-2.32*
			Internists and family practitioners per patient	0.0004	0.90
			R^2	0.14	
			DF	634	

*$p < .05$.
**$p < .01$.
***$p < .001$.

this simultaneous-equation approach. In the volume equation (on the left-hand side), a hospital's actual minus expected death rate for AMI patients has no effect on its volume. However, hospitals that (1) are larger than average, (2) have more appendectomies performed (perhaps indicating an active emergency service), and (3) have a Cardiac Care Unit (CCU) or Intensive Care Unit (ICU) receive more than the expected number of AMI patients. Medical school affiliation has no impact on volume, which is not surprising given the emergency nature of the admission. Finally, hospitals with more neighbors have higher volumes, probably because some of those neighboring hospitals do not receive emergency admissions, resulting in increased flows to hospitals that do. In the outcome equation, AMI volume has a significant negative effect, indicating better results. Mortality rates are significantly lower in the West, and, as LoGerfo and Shortell found, death rates are lower in hospitals with medical staff on the governing board. However, contrary to their findings, the ratio of primary care physicians per myocardial infarction patient is insignificant.

RESULTS FOR ALL DIAGNOSES AND PROCEDURES

Similar pairs of regressions have been estimated for each diagnosis and procedure under consideration. While the results for certain variables, such as geographic region and medical school affiliation, may be of interest to others, our focus is on the selective-referral and practice-makes-perfect hypotheses. (An appendix with the full set of regression results is available from the authors.) A significant negative coefficient on the volume variable in the equation explaining death rates will support the practice-makes-perfect hypothesis. A significant negative coefficient on the actual-expected death rate in the volume equation supports the selective-referral pattern hypothesis. Table 9 provides a summary of the findings for the two sets of coefficients for each diagnosis and procedure. While not all the results are as one would expect, there are important general patterns.

Procedures and diagnoses falling into Cell 2 exhibit high volumes resulting in lower death rates, but there is no measurable influence of outcomes on volume; that is, the practice-makes-perfect effect predominates. All four of these, acute myocardial infarction, stomach operations, intestinal operations, and cholecystectomy, are usually managed by the family physician or a general surgeon, perhaps in consultation with local cardiologists or gastroenterologists. There is little reason to seek outside specialty consultations, so referrals to centers with particularly good outcomes are unlikely.

Table 9: Pattern Effects

Practice-Makes-Perfect: *Effect of Volume on Death Rate**	*Selective-Referral Patterns: Effect of Death Rate on Volume**		
	Negative	*Insignificant*	*Positive*
Negative	Total hip replacement Hysterectomy 1	Acute myocardial infarction Cholecystectomy Stomach operations Intestinal operations 2	Respiratory distress syndrome 3
Insignificant	Aneurysm Fracture of femur Ulcer Transurethral prostatectomy Cardiac bypass graft 4	Cirrhosis Hernia repair 5	Appendectomy 6
Positive	7	8	Subarachnoid hemorrhage 9

Cell

1. Both "Practice-Makes-Perfect" and "Selective-Referral Patterns."
2. Primarily "Practice-Makes-Perfect."
3. Primarily "Practice-Makes-Perfect," Counterintuitive results for "Selective-Referral Patterns."
4. Primarily "Selective-Referral Patterns."
5. No clear relationship.
6. Counterintuitive results for "Selective-Referral Patterns."
9. Counterintuitive results for "Selective-Referral Patterns."

*Significance is measured at the .05 level.

For Cell 4, the selective-referral effect predominates. Volume has no effect on death rates, but hospitals with low death rates attract patients with abdominal aortic aneurysm, fracture of the femur, peptic ulcer disease, transurethral resection of the prostate, and coronary artery bypass graft. With the exception of ulcer, referral to a specialist is clearly indicated. The results for ulcer may reflect either referrals for refractory patients or small area-variation factors in which some physicians admit many patients with marginal indications and obtain good outcomes [19, 20]. (*Note*: these data predate the widespread use of cimetidine and the subsequent marked decline in hospitalization for ulcer.)

In Cell 1, both practice-makes-perfect and selective-referral patterns are observed. While this is plausible for total hip replacement, it appears unlikely that true selective referrals occur for hysterectomy. The latter results may be a reflection of locally high population-based hysterectomy rates in areas served by specific hospitals [19, 20]. In Cell 5, cirrhosis, hernia repair, and cardiac catheterization all show neither relationship. While there are certainly extensive referrals for catheterization, it is important to note that these figures represent referrals within the limited number of hospitals offering the procedure.

Three patient categories show a positive impact of death rate on volume. For two of these, respiratory distress syndrome and subarachnoid hemorrhage, mortality rates are very high. While it is possible that high-volume centers really do produce worse outcomes, it is more likely that the expected mortality measures were not sufficiently sensitive to measure risk differentials for referral centers. Furthermore, these two diagnoses have extraordinarily high transfer-out rates for low-volume hospitals and high transfer-in rates for high-volume hospitals. Since transferred patients are not tracked from one hospital to another, low-volume institutions will have an artificially depressed mortality rate. Certainly the discussion in the first part of this article supports the notion of selective referrals for these patients. The anomalous results for appendectomy may be due to a low death rate, so that a few hospitals with poor outcomes dominate the results. (It is also possible that these anomalous findings stem from the use of cholecystectomy as the "control" volume, rather than appendectomy. Appendectomy probably belongs in Cell 5, with no significant pattern in either direction.)

CONCLUSIONS AND IMPLICATIONS

Because the early papers on the relationship between volume and outcome used analogies to the "learning curve" relation found in the airframe industry, it became easy to view the data as if high volumes led to better outcomes [1, 11]. Such a relationship had strong policy implications favoring regionalization, which one might also want to encourage for other reasons, such as lower costs. One can ask, however: if the early discussions had used the restaurant analogy instead, arguing that high volumes are indicators of good quality, might the selective-referral hypothesis have become dominant? This latter hypothesis, too, is consistent with regionalization, but one must carefully identify the hospitals selected to be referral centers. Luft, Bunker, and Enthoven [1] noted the alternative policy implications, and Luft [17] used a simpler simultaneous-equation model to explore the alternative explanations. The purpose of this article has been to present a convincing case that both explanations are supported by the evidence, and that for particular diagnoses and procedures, different explanations seem more plausible. In presenting the case, we used both simple cross-tabulations and a simultaneous econometric test in an attempt to offer robust evidence to convince skeptics with various methodological backgrounds.

The argument is not conclusive. Not all of the patterns in the crosstabs are consistent with the regressions. The equations predicting outcomes have low-to-moderate explanatory power (R^2 = .02–.28) in contrast to the volume equations (R^2 = .13–.80). It is frustrating that we have not been able to identify structural measures, such as teaching status, to explain variations in outcomes across hospitals. Similarly, the risk-adjustment measures are relatively crude, and for some diagnoses (e.g., RDS) seem unable to identify hospitals with very high risk case mixes. It is also important to recognize that factors in addition to better outcomes may account for heavy patient concentrations. For example, inner-city hospitals may have a disproportionate share of patients with cirrhosis, teaching hospitals may serve as referral centers for subspecialty care, and hospitals in areas with many elderly may have a large number of patients with hip fractures. In multivariate analyses we have held constant many of these factors, but they seem to have little influence on differences in volumes. A confounding factor is that physician volume may be more important than hospital volume. Analyses of hospital and surgeon volumes (but not in a simultaneous-equation model) show that hospital effects continue to be important [21]. New data sets will be required to track physician volume across several hospitals, but it is unlikely that such data will markedly alter our conclusions.

In spite of these empirical shortcomings, the patterns across diagnoses and procedures are far from random. Patient categories exhibiting results consistent with selective referrals are, in general, more complex and more likely to be treated by a specialist than by a primary care physician or general surgeon. Likewise, those patient categories exhibiting results consistent with the practice-makes-perfect hypothesis are less likely to be referred. These findings are supported by such simple measures as the proportion of patients transferred and the risk differentials across hospitals.

What, then, are the implications for policy? If we could confidently reject alternatives to the practice-makes-perfect hypothesis, then those who favor concentrating patients in specific hospitals would not need to worry about how the hospitals are chosen. For example, California's Medicaid program now restricts its beneficiaries to a subset of the state's hospitals which were selected largely on the basis of per diem bids [15]. Insurers are developing preferred provider organizations with financial incentives for enrollees to use selected hospitals [14]. Again, selection is far more often based on price than on medical outcomes. In fact, if only the practice-makes-perfect hypothesis is applicable, then the selection criterion is irrelevant. Regionalization efforts will improve outcomes in the selected hospitals—and perhaps worsen them in those that lose patients.

However, if selective-referral patterns are an important explanation of the volume-outcome relation, selective contracting based merely on price will not necessarily channel patients to the better hospitals, nor will outcomes necessarily improve. Indeed, they may worsen markedly if the higher-quality settings are excluded and if increased volume in the selected hospitals does not improve outcomes sufficiently to offset the loss of the better providers. Furthermore, if some physicians feel strongly about quality differences influencing their referrals, restraints on their available choices may be resisted and such restraints could become factors in malpractice suits. These considerations suggest that efforts to concentrate patients in selected hospitals should be sensitive to the possibility that unmeasured, but nonetheless important, physician and hospital factors influence outcomes and that existing referral patterns may already reflect such factors. Policymakers and insurance companies must also be cautious, because the results suggest that the roles of practice, referral, and other factors vary across diagnoses and procedures.

ACKNOWLEDGMENTS

We are grateful to Ann Flood, Nancy Ramsay, and two anonymous referees for helpful comments.

REFERENCES

1. Luft, H., J. Bunker, and A. Enthoven. Should operations be regionalized? An empirical study of the relation between surgical volume and mortality. *New England Journal of Medicine* 301:1364–69, December 20, 1979.
2. Forrest, W. H., et al. Impact of Hospital Characteristics on Surgical Outcomes and Length of Stay. Final Report to National Center for Health Services Research, U.S. Department of Health and Human Services, (HRA) 230-75-0173, 1970.
3. Maerki, S. C., H. S. Luft, and S. S. Hunt. Selecting categories of patients for regionalization: Implications of the relationship between volume and outcome. *Medical Care* 24(2): 148–58, February 1986.
4. Shortell, S., and J. LoGerfo. Hospital medical staff organization and quality of care: Results for myocardial infarction and appendectomy. *Medical Care* 19(10):1041–53, October 1981.
5. Wolfe, R., et al. Mortality differences and speed of wound closure among specialized burn care facilities. *Journal of the American Medical Association* 250:763, 1983.
6. Riley, G., and J. Lubitz. *Outcomes of Surgery in the Medicare Population: The Relation of Surgical Volume and Other Factors to Mortality.* Baltimore, MD: Health Care Financing Administration, Office of Research and Demonstrations. Working Paper Series No. 84-6, August 1984.
7. Farber, B., D. Kaiser, and R. Wenzel. Relation between surgical volume and incidence of postoperative wound infection. *New England Journal of Medicine* 305(4):200–04, July 23, 1981.
8. Flood, A. B., W. R. Scott, and W. Ewy. Does practice make perfect? I: The relation between hospital volume and outcomes for selected diagnostic categories. *Medical Care* 22:98, February 1984.
9. Flood, A. B., W. R. Scott, and W. Ewy. Does practice make perfect? II: The relation between volume and outcomes and other hospital characteristics. *Medical Care* 22:115, February 1984.
10. Flood, A. B., W. R. Scott, and W. Ewy. Letter in reply to: Dranove, D. A comment on "Does practice make perfect?" *Medical Care* 22(10):967–69, October 1984.
11. Asher, H. Cost-Quantity Relationships in the Airframe Industry. Santa Monica, CA: The Rand Corporation, 1956.
12. Lieberman MB. The learning curve and pricing in the chemical processing industries. *Rand Journal of Economics* 15(2):213–28, Summer 1984.
13. Dranove, D. A comment on "Does practice make perfect?" *Medical Care* 22(10):967, October 1984.
14. Trauner, J. B. Preferred Provider Organizations: The California Experi-

ment. San Francisco: Institute for Health Policy Studies, University of California, San Francisco, 1983.

15. Iglehart, J. K. Cutting costs of health care for the poor in California: A two-year follow-up. *New England Journal of Medicine* 311(11):745–48, September 13, 1984.
16. Luft, H. S., and S. S. Hunt. Evaluating hospital quality through outcome statistics. *Journal of the American Medical Association* 255(20):2780–84, May 23-30, 1986.
17. Luft, H. S. The relation between surgical volume and mortality: An exploration of causal factors and alternative models. *Medical Care* 18(9):940–59, September 1980.
18. Luft, H. S., and C. Fox. Computation of Distances and Neighborhoods (Draft). San Francisco: Institute for Health Policy Studies, University of California, San Francisco, 1983.
19. Wennberg, J. E., and A. Gittelsohn. Variations in medical care among small areas. *Scientific American* 246:120, 1982.
20. Roos, N. Hysterectomy: Variations in rates across small areas and across physicians' practices. *American Journal of Public Health* 74(4):327, April 1984.
21. Hughes, R. G., S. S. Hunt, and H. S. Luft. The effects of low surgeon volume on hospital quality. *Medical Care* forthcoming, June 1987.