Chapter 8

Volume of Services in Hospitals or Performed by Physicians

CONTENTS

Figures

Tables

Chapter 8

Volume of Services in Hospitals or Performed by Physicians

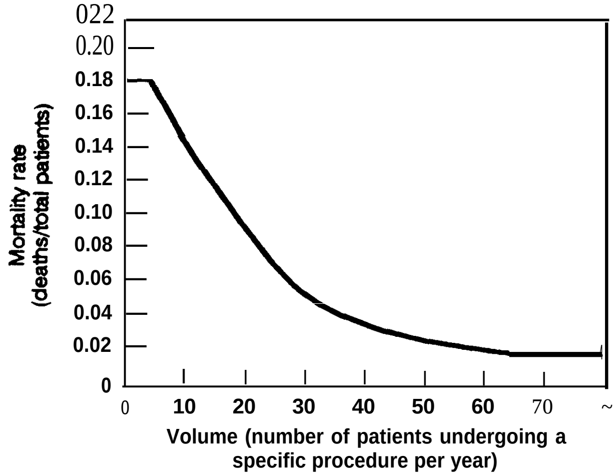
INTRODUCTION

There is a common notion that "practice makes perfect." In the medical care setting, this adage is often interpreted as "high-volume hospitals and physicians achieve better outcomes." The word "volume" in this context refers to the number of procedures or number of patients with the same diagnosis treated in a specific hospital or by a particular physician. For some procedures and diagnoses, better patient outcomes and lower inhospital mortality have been associated with higher volumes.

In its simplest form, the hypothesized relationship between volume and outcome may be displayed as a graph with volume (e.g., number of patients undergoing a specific procedure per year in a hospital) on the horizontal axis and outcomes (e.g., mortality rate) on the vertical axis. The graph in figure 8-1 shows high mortality in hospitals with low volumes and low mortality in hospitals with high volumes. The flattening of the curve at high-volume levels indicates that there is little additional reduction in mortality above a certain volume threshold.

It is important to limit the conclusions drawn from this graph. Even if a relationship is found between volume and outcome, it is inappropriate to conclude that increasing the volume in a hospital will improve outcomes or that reducing the volume will worsen outcomes. Conclusions cannot be drawn about how changes in volumes affect changes in outcomes, because most analyses use data from a cross section of hospitals observed at a point in time rather than data from the history of mortality and volume over time. Instead of causality from volume to outcome, there may be causality from outcome to volume; that is, medical providers with low mortality rates may attract higher volumes of patients. Another possibility is that some unmeasured factor may account for an observed relationship between volume and outcome. For example, high-volume hospitals or physicians may have relaxed admission criteria; these relaxed criteria, in turn, may mean that some of their patients are healthier and less likely to suffer adverse outcomes. In this case, both the higher volume and the better outcomes may be caused by the relaxed admission criteria.

Figure 8-1.-Hypothesized Relationship Between Volume and Outcome



SOURCE: Office of Technology Assessment, 1988

For OTA's review of the literature on the volume-outcome relationship, the abstracts of approximately 100 papers were read. Of the 50 articles that were thoroughly reviewed, 26 presented reportable findings.[2] Studies were included if they examined a sufficient number of hospitals (over 20) and cases to offer statistically valid volume-outcome results or if the study purported to ex-

[1]This chapter is based on a paper prepared for OTA by Harold S. Luft, Deborah W. Garnick, David Mark, Stephen J. McPhee, and Janice Tetreault (395).

[2]Additional technical information on the studies included in the literature review is available in the paper on volume prepared by Luft and colleagues (395).

amine the volume-outcome relationship. These studies pertain to both hospital and physician volume. Although most research relates to hospital volume, a growing body of literature focuses on the relative importance of physician volume in contrast or in addition to hospital volume.

This chapter examines the reliability, validity, and feasibility of using the volume-outcome relationship as an indicator of the quality of medical care, and explores the issues of causality as well as other relevant conceptual and methodological issues. How volume data might be used by consumers in choosing hospitals and physicians is discussed, and further necessary research is outlined.

RELIABILITY OF THE INDICATOR

Information on the volume of procedures and diagnoses and on inhospital mortality is routinely available from two sources: hospital discharge abstracts and insurance claims. There are several problems with the reliability of hospital discharge abstract data. Errors can occur at different points during the data collection process: in recording the patient's diagnosis or procedure onto the medical chart, in the translation of the chart onto discharge abstract forms, or in the transformation of discharge abstract forms into large-scale computerized data systems. Several studies of the accuracy of hospital abstracting suggest a high error rate (450,532). Moreover, inaccuracy in the data may be the result of random errors, such as misapplication of coding rules or the selection of vague diagnosis codes, or maybe the result of purposeful misspecification of a patient's principal diagnosis in order to achieve an optimal diagnosis-related group (DRG) for Medicare payment purposes. Recently, a reabstracting study noted that incorrect DRGs were originally assigned 20.8 percent of the time in 1984-85 and that 61.7 percent of these errors benefited the hospital (304).

Insurance claims data—especially non-Medicare insurance claims data—usually include less information about diagnoses than do routinely collected hospital discharge abstract data. Moreover, coding problems in the case of claims data may be worse than those in the case of hospital discharge abstracts. The problem is especially acute for diagnoses; procedures are generally well coded (131).

The pertinent question here, however, is not whether coding errors occur, but how such errors affect volume-outcome studies. The miscoding of a diagnosis or procedure may cause undercounts or overcounts of the number of patients in certain categories. Many of the diagnoses and procedures that have been studied in the volume-outcome literature are so important to a patient's hospitalization and the categories are so broad, however, that miscounts of patients are probably not an important concern. Total hip replacement, for example, would be unlikely to be overlooked. Moreover, in many studies, volume is specified as a series of categories (e.g., high, medium, and low), so a small amount of random undercounting or overcounting is not crucial. Miscoding of patients' simultaneously existing illnesses (comorbidities) may be a problem in case-mix adjustments to reflect patient differences. The problems in adjusting for patient differences in the analysis of volume are similar to those present in the analysis of hospital-specific mortality outcomes (see ch. 4).

Volume-outcome studies are generally cross-sectional, and changes in the accuracy of data over time are less important than systematic differences across hospitals. When analyses are focused on individual hospitals, the reliability of data is an important concern, because misclassification could result in the mislabeling of a hospital as a good- or poor-quality provider. When the investigation concerns the identification and exploration of the hypothesized relationship between volume and outcome, the reliability of data is less of a key concern. Suppose there are random errors across hospitals in the coding of diagnoses. Such errors will affect the precision with which relationships are estimated, but if the errors are uncorrelated with volume, the volume-outcome effect will not be altered.

VALIDITY OF THE INDICATOR

Table 8-1 presents a summary listing of the 15 procedures and diagnoses investigated in the 26 studies used for the analysis in this chapter. The studies are grouped in the left hand column by research team and by the publication date of the first article by the team (e.g., all three studies by Kelly and her colleagues are shown together). To check which authors studied a particular diagnosis or procedure, read down the column for a given procedure or diagnosis.

Of the 15 procedures and diagnoses investigated in the 26 studies, 13 are surgical procedures. Only 2 are medical diagnoses: acute myocardial infarction ("heart attack") and newborn diseases. The study of surgical procedures is easier than the study of medical diagnoses for several reasons. First, surgical procedures are generally well identified and coded both on hospital discharge abstracts and insurance claims. The occurrence of an operation is rarely in dispute, even though the choice of procedure or necessity for it may be questioned by various physicians. s The determination of some diagnoses, on the other hand, is often quite difficult; comparably trained clinicians may disagree on an individual patient's diagnosis.

Second, although severity of illness may vary with both surgically treated patients and medically treated patients, it is less likely to be a major source of bias in volume-outcome studies of patients treated surgically. Surgery is usually used to increase longevity or to correct a problem that interferes with the quality of a person's life but is not immediately life-threatening. Thus, a surgically treated patient is often in reasonably good health on admission to the hospital, and short-term mortality is more likely to reflect the effects of treatment than to reflect the patient's initial health status. In medical admissions, on the other hand, there is greater variation in the complexity of cases, and a patient's health status on admission may be a more important determinant of short-term outcomes than the quality of care rendered is. Thus, the paucity of good measures of patients' severity of illness probably has a greater impact on studies involving medical admissions than studies involving surgical admissions.

[3]In some cases, there may be miscoding of which procedure occurred, for example, revision of total hip versus total hip replacement.

Measures of Volume and Outcome

Volume is measured in several ways in the 26 studies reviewed by OTA:

- categorical variables (e.g., low-and high-volume groups, or a four-or five-category classification),
- a continuous variable (e.g., a count of number of patients, which allows for a linear relation),
- volume and volume squared (which allows for either linear or "U'' -shaped curves), or
- log of volume (which allows for a stronger effect at low volumes and progressively weaker effects at higher volumes).

Most of the studies measure volume for a single year, although some studies use other periods. One study uses a hybrid: the proportion of patients in a hospital (a continuous measure) treated by surgeons with low volumes (a dichotomous variable) (307).

Four measures of patient outcomes are used in the 26 studies:

- inhospital mortality,
- mortality within a fixed period of time,
- complications or health status measures, and
- long hospital stays as a proxy for complications.

The use of mortality as an outcome measure of quality has some limitations (see ch. 4). For some procedures and diagnoses, mortality is so rare an event that it is difficult to determine whether an occasional death indicates a pattern of poor quality or a chance occurrence. Important biases may also be introduced because discharge policies controlled by hospitals can affect inpatient mortality rates. In hospitals that transfer patients with severe complications to other, more appropriate facilities, such as regional tertiary hospitals, there are likely to be lower mor-

Table 8“1.—Studies Reviewed by OTA on the Relationship Between Volume and Outcome for Specific Diagnoses and Procedures

	Abdominal aortic aneurysm	Acute myocardial infarction	Appendectomy	Bilary tract surgery	Cardiac catheterization	Coronary artery bypass graft surgery	Femur fracture	Hernia	Hysterectomy	Intestinal operation	Newborn diseases	Prostatectomy	Stomach operation	Total hip replacement	Vascular surgery
1. Adams, et al., 1973 (5)					HVID	‘									
2. Wilhams, 1979 (702)															
3. Luff, et al., 1979[b] (394)	HV/D			HV/D		HV/D				HV/D					
4. Luft, 1980[c] (393)	HV/D			HVID		HVID				HV/D					
5 Maerki, et al., 1986[d] (402)	HV/D	HV/D	HV/D	HV/D	HV/D	HV/D	HV/D	HV/D	HVID	H V / D	*				
6. Luff and Hunt, 1986 (396)					HV/D										
7. Luft, et al., 1987[e] (397)	HV/D	HV/D	HV/D	HV/D	HV/D	HV/D	HV/D	HV/D	HV/D	HV/D	HV/D	HV/D	HV/D	HV/D	
8. Hughes, et al., 1987 (307) :			HV, PVID, L	HV, PVID, L	HV, PVID, L	HV, PV/D, L		HV, PV/D, L	HV, PV/D, L	HV, PVID, L		HV, PV/D, L	HV, PVID, L	HV, PV/D, L	
9. Hughes, et al., in press (306)							HV/D, L								
10. Pilcher, et al., 1980 (488)	HV,PV/D														
11. Farber, et al., 1981’ (203)			HV/M					HV/M	HV/M	HV/M					
12. Shorten and LoGerfo, 1981 (571)’		PV/D	PV/M												
13, Hertzer, et al., 1984 (295)															PV/D,M
14 Flood, et al., 19849 (217)	HV/D			HV/D			HVID			HV/D			HV/D		
15. Rosenblatt, et al., 1985 (538)											HV/D				
16, Riley and Lubitz, 1985 (520)				HVID		HV/D	HVID	HV/D		HV/D		HV/D		HV/D	
17. Kempczmskl, et al., 1986 (349)															PV/D,M
18 Sloan, et al., 1986[h] (582)						HV/D			HVID					HV/D	
19, Kelly and Helhnger, 1986 (347) : :	HV, PV/D									HV, PV/D			HV, PVID		
20. Kelly and Hellinger, 1987 (348)		HV, PV/D			HV, PV/D	HV, PV/D									
21. Kelly, forthcoming’ (346)				HV/D	HV/D	HVID									
22. Roos, et al., 1986 (531)				HV, PVf R					HV, PV/R			HV, PV/R			
23. Roos, et al., 1987 (533)				HV/R					HVIR						
24. Wennberg, et al., 1987 (697)												HV/D, R			
25. Showstack, et al., 1987 (573)						**HV/D, L**									
26. Fowles, et al., 1987 (227)														HV, PV/M, D	

Abbreviations HV = hospital volume, PV = physlclan volume, D = death, L = long length of hospital stay, M = morbldlty, R = readmmon
asfudies are ordered by research t~m and date of first publlcatlorl by the **feam Numbers m parentheses refer to numbered entnes m the reference Ilst at fhe end of h's ‘epon**

[b]Luff, et al (1979), alSO studied open-heart surgery
[c]Luff, et al (1980), also studied open-heart surgery
dMaerk, et al (1986), also studied clrrhosls, peptic ulcer, subarachnold hemorrhage. and tonsillectomy

~Luft, et al (1987), also studied clrrhows, subarachnold hemorrhage, and peptic ulcer
Farber, et al (1981). also studied lammecfomy and cesarean secflon
gFlood, et al (1984), also studied amputation of lower Ilmb, nonsurgical gallbladder dlagnosls, and nonsurgical ulCM dlagnosls

‘Sloan, et al (1986). also studied morbid obesity surgery, mastectomy, nephrectomy, and spinal fusion
IKelly (fofihcomlng) also studied atherosclerow, cranial wry, dla~tes. and hyperfens’on

SOURCE Ofhce of Technology Assessment, 1988

tality rates. In hospitals with longer average stays, there is a greater chance of observing a death. Suppose, for example, that one hospital typically keeps patients for 10 days after a certain surgical procedure, while another hospital works to get patients on their feet and discharges them after a week. If a certain fraction of patients from each hospital experiences a fatal heart attack on the 8th to 10th days after surgery, these deaths will be counted in the inhospital mortality rate of only the first hospital. Because of these biases, some researchers calculate mortality rates with respect to a fixed window, such as 30 days, after admission (643).

Complications and other measures of patients' health status are less objectively measured than mortality. In some instances, a clearly identified procedure, such as a reoperation, indicates a poor outcome. Other measures, such as surgical wound infections, are less reliably coded across hospitals (see ch.5).

One final measure of quality is even further removed from a direct measure of outcome. Luft and his colleagues use the proportion of patients that stay a very long time in the hospital as a proxy for complication rates (306,307,573). They argue that if one chooses a length of stay exceeded by only 10 percent of all patients, then a hospital with far more than 10 percent of its patients staying that long or longer may be experiencing poor outcomes. Although this argument is plausible, it has not been validated by determining whether those patients with very long hospital stays truly have complications, or stay longer, for example, because nursing home beds are scarce.

Differences in Patient Characteristics

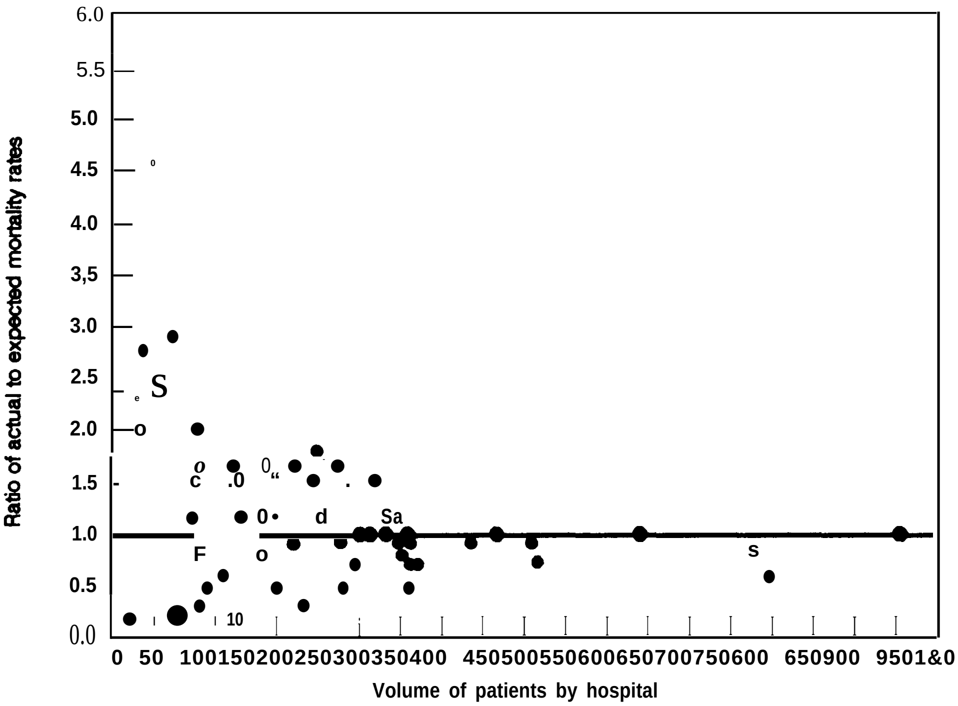
A major problem in analysis of the volume-outcome relationship is the potentially confounding effect of differences in patient characteristics. Every patient is different, and individual factors strongly influence outcomes. Even if these patient differences are random, the estimation of a volume-outcome relation will be made more difficult because of the "noise" due to these random effects. This point is illustrated in figure 8-2 which plots the inpatient mortality rates for patients undergoing coronary artery bypass graft (CABG) surgery in 78 California hospitals in 1983 (574). Although there generally appear to be lower rates of poor outcomes at higher volumes (a negative linear relationship), there is substantial variation among hospitals at given volume levels—variation due in part to patient-related factors.

The crucial question is whether more or less severely ill patients are consistently admitted to high-volume hospitals. If they are, an observed association between outcome and volume could be due entirely to patient mix. The true answer to this question would be found by random assignment of large numbers of patients to institutions with varying volume levels. Random assignment, with sufficiently large numbers of patients, would reduce to insignificance the likelihood that patient-related factors account for the observed differences in outcomes. Unfortunately, since such an experiment would be enormously expensive and impossible because of ethical considerations, one is left with attempts to control for patients' differences by various statistical means.

There are two general approaches to dealing with differences in patient mix across hospitals. The first is to specify the procedure or diagnosis for study as carefully and narrowly as possible. The intent of this approach is to set patient selection criteria that result in a homogeneous group of patients. For example, patients undergoing CABG surgery who also have heart valve surgery have mortality rates about three times as high as those of patients undergoing CABG surgery only (573). Since some hospitals may specialize in uncomplicated CABG surgery while others have a large share of patients also requiring valve surgery, results may be biased unless one focuses on patients with CABG surgery only.

The second approach, which can be combined with the first, is to include variables in the analysis that may capture risk differences among the patients included in the study. In theory, each of these additional variables could be used to further stratify the study population of patients, but this approach is limited by an ever shrinking sample size. In many studies, therefore, patient selection criteria are combined with statistical controls. The patient's age, race, and sex are classic variables used in analyses. Transfer from another hos-

Flgure 8-2.-Ratio of Actual to Expectsd Mortaltty Rates by Volume of Patients Undergoing Corona ry Art- Bypass Graft Surgery In Caltfornla, 1983



KEY:
● = 1 hospital
O = 2 hospitals

SOURCE: JA Showstack, KE Rosenfeld, DW Garnick, et al., Institute for Health Policy Studies, University of California, unpublished data, San Francisco, 1987

pital is often a powerful indicator of a patient at higher risk of a poor outcome (393). Counts of the number of secondary diagnoses or procedures or the presence of specific diagnoses or procedures also are used (394,573,582). In some instances, diagnostic information is combined to form a disease "stage" indicative of the severity of the principal diagnosis (346,347,348).

The problem of differences among patients has been highlighted in the literature on using mortality data to evaluate hospital performance (78, 189). To some extent, the problem is more severe if the focus is on studying individual hospitals rather than on studying the hypothesized relationship between volume and mortality. If a specific hospital is identified as having a significantly above average mortality rate, the hospital administration is likely to claim that unmeasured differences in patient mix account for the observed results. Upon careful examination of the medical records, one may find that some patients entering the hospital with severe problems do account for an elevated mortality rate (see ch. 4). Precisely what clinical characteristics, if any, are similarly correlated with volume is not clear.

Research Findings

Statistical methods used in the volume-outcome studies listed in table 8-1 range from simple comparisons of high- and low-volume groups to fairly sophisticated causal models. Regression models were commonly used because they can include a large number of patient and/or hospital variables as explanatory factors. In some cases, logistic models were used to account explicitly for the 0,1 nature of patient mortality. Three papers used simultaneous equation models to estimate both the influence of volume on outcomes and the influence of outcomes on volume (307,393,397).

Some researchers used the patient as the unit of observation in a regression model to "predict" the patient's outcome. These researchers included as many patient risk factors as possible as well as variables indicating the number of patients with the procedure or diagnosis in the hospital per year. Patient-level regressions typically produced very low R-squares, indicating a low ability to predict whether an individual patient will live or die, even though many of the variables, such as volume, may be highly significant. Other researchers have argued that the focus of volume-outcome studies is on the average performance of hospitals at different volume levels, so the number of observations should be the number of hospitals rather than the number of patients undergoing procedures (701). These researchers estimated models at the hospital level that include the proportion of patients with each risk factor to predict a composite expected poor outcome rate based upon patient mix in the hospital.

It is difficult to combine or compare results across studies in a formal manner because of the differences in methods. For example, it is impossible to compare directly the findings of one study that simply contrasts mortality rates for hospitals with volumes above or below an arbitrary cutoff with another study that estimates the influence of the log of volume on outcomes while controlling for numerous hospital characteristics and referral effects.

To overcome this problem, OTA categorized the results of each study in terms of the implicit shape of the volume-outcome relationship "curve." Tables 8-2 and 8-3 summarize the results of this categorization for hospital and physician volume, respectively. Potential categories illustrated by the curves in the far left column of each table are as follows:

1. dichotomous results, with volume grouped into two categories and results showing lower rates of poor outcomes in high-volume settings;
2. a negative linear relationship, also showing lower rates of poor outcomes in high-volume settings;
3. a "U-shaped" relationship showing higher rates of poor outcomes at lower volumes, lower rates of poor outcomes at intermediate volumes, and higher rates of poor outcomes at higher volumes;
4. an inverse logarithmic form, with large reductions in the rates of poor outcomes as low volumes increase and a relative flattening at high volumes;
5. a "flat" curve, indicating no significant relationship; and
6. a positive linear relationship, with higher rates of poor outcomes at higher volumes.[4]

Tables 8-2 and 8-3 should be read along with table 8-1, which lists the 26 studies included in OTA's literature review along with the diagnoses and procedures examined in each study. In table 8-2, for example, results for abdominal aortic aneurysm are shown in the first column. The first study in that column, number 14, refers to study number 14—Flood, et al., 1984—listed in table 8-1.

When one of the 26 studies includes two methods (e. g., dichotomous and continuous-volume variables) or differentiates between two subcategories of a procedure (e.g., ruptured aneurysm surgery and elective aneurysm surgery), the results are counted separately. As an example, table 8-2 shows that seven studies addressed hospital volume and abdominal aortic aneurysm. Using regression analysis, Luft and his colleagues found an insignificant (flat) relationship between volume and outcome for this procedure (study 7c); however, using volume categories (study 7a), the same authors found a negative linear relationship between volume and outcome for this procedure (397). Table 8-2 shows both these results.

When reviewing a set of findings such as these, which are relatively thin but nonetheless cover many procedures and diagnoses, one is torn between a "lumping" approach to provide a gestalt and a "splitting" approach to explain differences. Several points can be highlighted for specific diagnoses or procedures. For biliary tract surgery, it is important to distinguish the type of surgery, because the volume-outcome relationship may be valid only for more complex surgery that com-

[4]One set of findings by Sloan, Perrin, and Valvona (582) approximates a backwards "C" and is not classified in this schema. These investigators' other findings fit a "U" -shaped pattern and are included in the table.

Table 6-2.-The Hospital-Volume/Outcome Relationship: Summary of Research Findings From Studies Reviewed by OTA on Specific Diagnoses or Procedures*

Shape of the curve	Mctomine Sortlc aneurysm	Acute nyocard!a mfarotm	Appendec-tomy	Biliary tract surgery	Cardiac catheter-ization	Coronary artery by ~ surge~	Femur fracture	Hernia	-fy3terec-tomy	Intestinal operation	Newborn dwmses	Proatatec-hrny	Stomach operation	Total hip ment	vascular surgery
	14			14,23					23 -	14 t		-	t -	?	
2.	5,7a, IOb	7a	5,7a	5,7a	1,5	'a	5,7a						5,7a		
3.		5			7a	5		5			2,5			5,18	
4.	3,4,19	7C	8,11	3e,3f,4e, 4f,4g,7c, 11	6,8,20, 21	3,4,8, 16:20, 25425j	91,9m	7a,8,11	,8,20, 1	3,4,8, 16,20, 25i,25j	7a,7c			3,4,7C, 8,16r	4
5.	7c,lod	20	7C	~,8J6,		7c,21 , 25k +	5,7a,7c, 14,16	7c, 16	T						
6.						t					15				

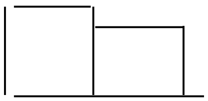
● Note: The curves m this table illustrate the follwmg relationships between volume and outcome:
1 Volume grouped mto two categories, with lwer rates of poor outcomes at higher volumes;
2 Downward-slooinra hne (necmtive hnear relationship), with lower rates of Poor outcomes at higher volumes,
3 U-shaped cuwe, ~ith htbh~ rates of poor outcomes at lwer volumes, lwer rates of poor o&omes in mtermadiate volume ranges, and higher rates of poor outcomes at higher volumes;
4 L-shaped curve (mvem-logarrfhmlc form), wtih large raduchons m rates of poor outcomes as volume increases from lw levels ad httle ch&ge at higher volumes,
5 Flat, mchcadmg no slgmficanl reiatlonshlp between volume and outcome, and
6 Upward-sloping hne (posmve hnear relationship), with higher rates of poor outcomes at higher volumes

The numbers m the entnesforthistable refer to the numbered references m table 8-1 The meamngs of the tatters m the entrms are as follows:

a Analysis using volume categorms
b Ruptured aneurysm surgery
c Analyws using regressions
d Electlve aneurysm surge~
e Cholecyatectomy w~h common bile duct exploration
f Other blliary tract surge~
9 Cholecystectomy alone
h Nonscheduled CASG, death outcome
ı Scheduled CASG, poor outcome
ı Nonscheduled CASG, poor outcome
k Scheduled CABG, death outcome
I Femur fracture, death outcome
m Femur fracture, poor outcome
n Vagotomy andlor pyloropldy for duodenal ulcer
o Vagotomy, all
p Stomach operations, cancer dlagnosm
~ Stomach operations, ulcer dmgnows
r Other hlp arfhoplasty
s Total hip replacement
t Total hip replacement, death outcome
u Total h!p replacement, major comphcatlons

SOURCE: Offlce of Technology Assessment, 19SS

Table 8-3.-The Physician-Volume/Outcome Relationship: Summary of Research Findings From Studies Reviewed by OTA on Specific Diagnoses of Procedures*

Shape of the curve	Abdominal aortic meurysm	Acute myocardial infar~on	Appendectomy	Biliary tract surge~	Cardiac catheterization	Coronafy artery bypass surge~	Hernia	Hysterec. tomy	Intestinal Dperation	pros~tectomy	Stomach operation	Total hip replacement	Vascular surge~
1. 		12	8	22	8		8			8		26e	
2. 	10a												13g
3. 													
4		20											
5.	10b,19			8,23	20	20		8,22,23	20		8,22,23	8,19	13h,13i, 13j, 17
6. 						8							

'Note: The curves m this table Nustrate the following relationships between volume and outcome:

1 Volume grouped mto two categories, wrth lower rates of poor outcomes at higher volumes,
2 Oownwa~d-sloprng line (negaf&e hnear relationship), wrfh lower rates of poooufcomes at higher volumes,
3 U-shaoed curve. with higher rates of ooor outcomes at lower volumes, lower rates of mor outcomes m mtermedlate vOhJme ranm?s, and hmher rates of poor outcomes at higher volumes;
4 L-sh&t curve iinverse"iogartihmic ~orm), wrth large reductions m ratea of poor out&mes as volume mcreasas from low levels &d Irttle ch-mge at higher volume%
5 Flat, mdicatmg no significant relationship between volume and outcome, and
6 Upward-sloping line (pxrtrve knew relatlonshlp), with higher rates of poor outcomes at higher volumes

The numbers m the entries for this table refer to the numbered rderences m table 8-1 The meanmgs of the letters m the entnes areas folbws:

a Electwe aneurysm surgery
b Ruptured aneurysm surgery
c Stomach operations, cancer diagnows
d Stomach operations, ulcer dlagnosm
e Poetoperative deaths
f Mqor comphcatlons
g Femoral pophteal bypass, ampufatlon outcome
h Carotid endarterectomy
i Aortofemoral bypass
j Femoral pophteal bypass, death outcome

SOURCE: Office of Technology Assessment, 19SS bines cholecystectomy with common bile duct exploration and other operations on the biliary tract. For CABG surgery, the hospital volume-mortality relationship may be driven primarily by emergency (study 25h) rather than scheduled patients (study 25k)(573). Similarly, Pilcher, et al., showed a volume-outcome relationship for ruptured (study l0b) but not elective aneurysm surgery (study 10d)(488).

Examining tables 8-2 and 8-3 overall makes it clear that a far greater number of available studies relate to hospital volume (table 8-2) than relate to physician volume (table 8-3); furthermore, many more studies of physician volume than of hospital volume found no relationship between volume and outcome. This pattern probably reflects three factors. First, physician volume data have been more difficult to obtain than hospital volume, so there have been more opportunities to undertake hospital studies. Second, even when data on physician volume have been available, it has been difficult to identify which physician is truly responsible for a patient when several specialists and consultants have been involved in a case. Third, some of the apparently inconsistent findings for physician volume may be due to methodological differences. Kelly and Hellinger (study 20), for example, found no surgeon volume-outcome relationship for cardiac catheterization when low-volume providers were omitted from their study. Hughes and his colleagues, focusing on low-volume surgeons, however, found worse outcomes associated with low-volume surgeons (study 8).

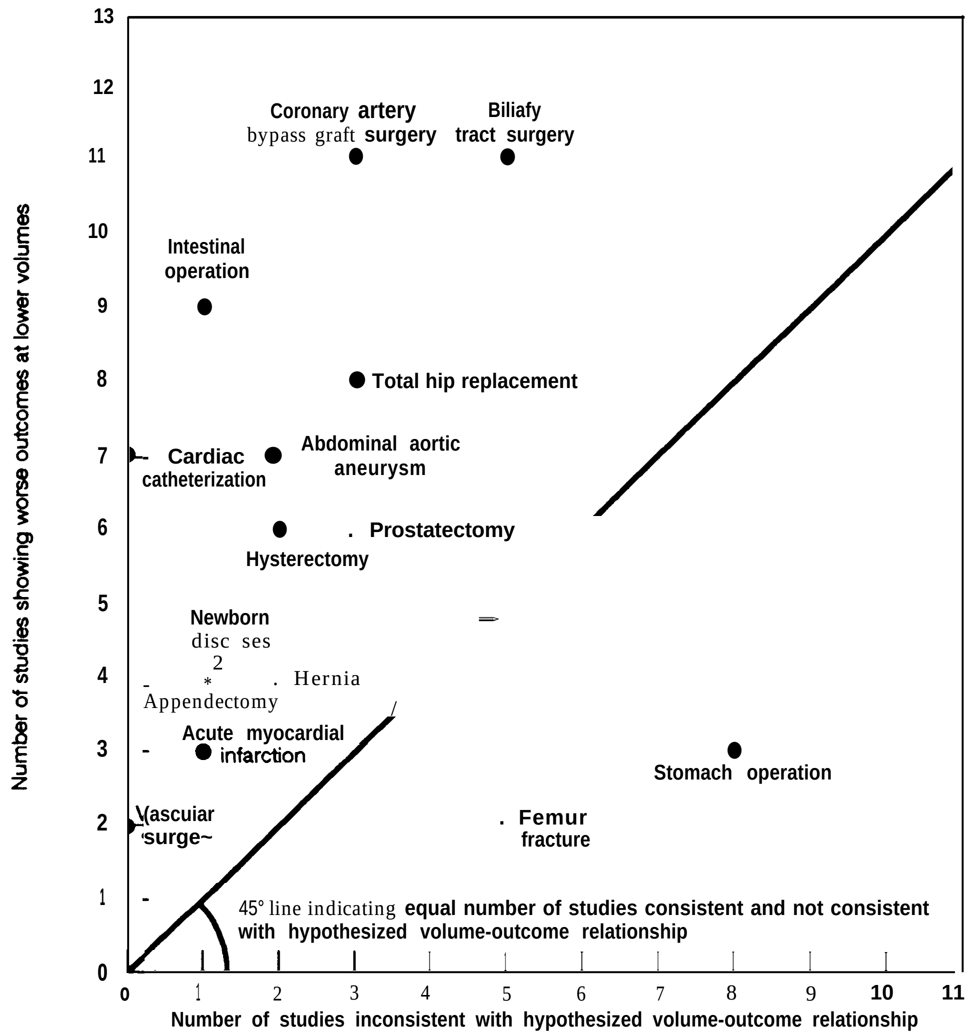
Without exception, none of the regression studies explicitly test a log versus a “U’’ -shaped curve, and there is little evidence of many observations on the upward sloping part of the “U.” Therefore, it is possible to lump the first four types of findings as all supporting the notion that worse outcomes tend to occur in low-volume settings. (This is not necessarily the same as saying that more is better.)

Two types of results across procedures and diagnoses are summarized in figure 8-3: findings that are consistent with the hypothesis that worse outcomes occur at lower volumes and findings that are inconsistent with that hypothesis. For hospital volumes of abdominal aortic aneurysm, for example, there are seven studies indicating worse outcomes at lower volumes (Y axis) and two studies showing no relationship between volume and outcome (X axis). For each of the 13 diagnoses and procedures in the upper left half of figure 8-3, there are more studies showing worse outcomes at lower volumes than studies showing inconsistent findings with regard to the hypothesized volume-outcome relationship. Worse outcomes are demonstrated at lower volumes in 11 of 14 studies of CABG, in 9 of 10 studies of intestinal operations, in 8 of 11 studies of total hip replacement, and in all 7 studies of cardiac catheterization. Only for the two procedures in the lower right half of figure 8-3 (femur fracture and stomach operation) are there more findings of no effect of volume on outcome than of worse outcomes at lower volumes.

Although detailed analyses of the methods used by each study reviewed by OTA are necessary to understand why results differ for a single diagnosis or procedure, several important factors help explain inconsistencies across studies: 1) physician vs. hospital volume, 2) causal linkages from volume to outcome or outcome to volume, and 3) the problem of detecting an effect if the rate of poor outcomes is low and the sample size is small.

Relatively little work has been done to distinguish various causal linkages in the volume-outcome relationship. Hospitals with high volumes are often institutions in which physicians have high volumes, and it maybe physician volume that truly matters. Therefore, it is crucial to distinguish between effects due to hospitals and effects due to physicians. Of the 124 findings concerning the effect of hospital volume on outcomes, 100 pertained to hospital volume without including physician volume, and 24 pertained to hospital volume and physician volume concurrently. Almost three-quarters of the 100 studies of hospital volume alone indicated a hospital effect, while only about half of the studies testing hospital and physician effects concurrently indicated a hospital effect. It appears, therefore, that in some instances, a measured hospital effect may be substituting for an untested physician effect. Alternatively, the high collinearity between physician and hospital volume may make it impossi-

Figure 8-3.-Number of Studies Reviewed by OTA Showing Either Worse Outcomes at Low Hospitsi Voiume or No Effect, by Diagnosis or Procedure



SOURCE: Office of Technology Assessment, 19SS

ble to detect true effects. Given the paucity of physician volume studies, one should reserve final judgment on this issue.

The uncertainty in this area reflects our limited understanding of the underlying reasons for the observed relationship between volume and outcome. The "practice-makes-perfect" explanation of the volume-outcome relationship rests on the general notion that increased experience results in more finely developed skills and, therefore, in better outcomes. The surgeon who consistently performs many units of a specific procedure will maintain, or continue to improve, his or her skills, while the surgeon who performs few procedures will become progressively less proficient. Similarly, 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 also make it possible for hospitals to purchase specialized equipment for such patients (217). Determining why outcomes for patients undergoing specific surgical procedures are related to volume requires extensive reviews of patients' medical charts from a large number of hospitals across a large number of procedures and diagnoses, because detailed data are unavailable from discharge data sets. For some procedures, problems in surgical technique may be the crucial factor, while for other procedures, inadequate postoperative monitoring may cause poor outcomes.

Even if physician volume is most important, hospital volume is likely to play a role. For example, a hospital with several high-volume and several low-volume surgeons may develop monitoring methods and standard procedures for the staff that catch errors and institute corrective actions. Thus, a low-volume surgeon maybe "protected" in a high-volume hospital. Likewise, a surgeon with a high volume across several institutions but low volumes in each may achieve good results. The empirical testing of such hypothetical relationships is difficult because of the need to track data on the same physicians across hospitals.

Volume may not matter at all, but instead may serve as a marker for hospitals or physicians with special skills whose better-than-average performance attracts a disproportionate share of the referrals. This "selective-referral" hypothesis holds that any inverse relationship between volume and outcome arises from the attraction of more patients to physicians and hospitals with better outcomes. The idea that patients in some instances may look for hospitals or physicians with the best results seems implausible to some, who claim that the variation in mortality by disease or procedure is too small to influence patients' choice (218). If complications are correlated with mortality, however, variations in outcomes may be large enough to be noticed by patients' primary physicians who choose specialists for referral. Although it is difficult to identify an individual hospital or physician as having significantly worse than average death rates (396), referral patterns may be based on a simpler set of decision rules. If primary physicians switch referrals after even one "bad outcome, " patients eventually are directed away from providers whose outcomes are worse. than average.

Furthermore, even if the majority of patients go to the nearest hospital or otherwise make decisions independently 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. As a result, hospitals with better outcomes would 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.

Another principally empirical objection to the selective-referral hypothesis is that some studies show little relationship between outcomes and hospital characteristics traditionally considered to be markers of good performance, such as teaching status or board certification of physicians (217,393). However, these measures are rather blunt and invalidated indicators of special expertise. It is common for a teaching hospital to be outstanding in the treatment of one diagnosis or procedure (e.g., cardiovascular surgery) but not to be particularly distinguished in another (e.g., neurosurgery).

When one attempts to test in a simultaneous-equation model both the effects of volume on outcomes and the effect of outcomes on volume, one may observe statistically significant effects for only one causal path. Even if the results indicate just an effect of outcome on volume in such a model, a simple test of volume as a function of outcome alone would probably show a relationship. There is not yet enough work to clearly indicate which causal paths are truly valid.

In designing an experiment, one should undertake a power test (ideally ahead of time) to determine the likelihood of detecting an effect if one truly exists. A power test is based upon the overall likelihood of the outcome's being measured and the sample size. There are substantial differences across studies in the number of patients involved and the average poor outcome (or mortality) rate,

To provide a sense of the issue at hand, consider the research findings from the 11 studies that reported on the hospital-volume/outcome relationship for the total hip replacement procedure. Eight studies showed a relationship between worse outcomes and low-volume hospitals, while three studies found no effect of volume on outcome (see table 8-2 and figure 8-3). The three studies that showed no effect had smaller sample sizes—under 1,500 patients in two studies and under 10,000 patients in the other study—than the sample sizes of from 13,700 to 33,000 patients in the eight studies that did find an effect. The three studies that had findings inconsistent with the hypothesized volume-outcome relationship probably had insufficient power to detect an effect unless it was very large. The mixed results for total hip replacement are not surprising given the design of the studies.

In summary, the available studies reviewed by OTA provide rather substantial evidence that worse outcomes occur at lower volumes for most of the procedures and diagnoses that have been studied. However, the volume-outcome relationship is not universal. For stomach operations and fractures of the femur, the evidence of a relationship is quite mixed, with the majority of studies indicating that volume has no effect on outcome. With the exception of the findings for stomach operations and femur fractures, all the other findings that suggest the lack of a relationship between volume and outcome either have low statistical power; are part of larger analyses in which a physician volume effect is found; or suggest a causal linkage from outcome to volume. Thus, although a relationship often exists, there is not yet enough evidence to distinguish effects due to physicians from effects due to hospitals or to have much confidence in the relative importance of the causal linkages.

FEASIBILITY OF USING THE INDICATOR

As has been discussed, there is frequently a relationship between volume and outcome. The general pattern is that better patient outcomes are associated with higher inhospital volumes. However, because there is hardly ever a perfect relationship, there are always some low-volume hospitals with apparently good outcomes and some high-volume ones with poor outcomes. This situation raises the obvious question, "How useful is volume as an indicator of the quality of care?" Since mortality data on Medicare patients are routinely available, why bother with volume data?

There will always be some chance component to a hospital's reported death rate in any single year, even after all adjustments for patient characteristics have been included. Various statistical calculations are designed to provide measures of this chance component and thus the degree of confidence one should have in the observed results for a particular hospital. It is inherent in the nature of small samples that one must expect much more variability in observed outcomes in hospitals with low volumes. One death among 10 or 20 patients may produce a mortality rate well above the average, but it is likely to be a chance occurrence. Similarly, even if the true or long-run mortality rate for that hospital is worse than average, with few patients in any particular year, there will often be years in which there are no deaths. To get a better estimate of the true performance of the outcomes in a low-volume hospital, one might aggregate data over several years, if they are available. Unfortunately, this technique makes it impossible to determine whether outcomes are improving or getting worse.

Combining data on volume and outcome is an alternative way of organizing a given amount of data to reduce the influence of chance and provide useful information. By aggregating data across hospitals within volume categories or using a regression to smooth out hospital-specific variability, the volume-outcome studies provide much more stable estimates of the performance of a class of hospitals. Although average results for all low-volume hospitals may not apply to a particular low-volume hospital, it is important to remember that, because of chance variability, last year's mortality rate for a particular hospital is not a very reliable indicator either. The two pieces of information, however, may be used together to guide a decision about a particular hospital.

The situation is different for high-volume hospitals, because the role of chance is smaller the larger the number of patients. Of course, hospital-specific mortality results will still be sensitive to unmeasured differences in patient characteristics that may not be adequately captured in the available data. If a high-volume hospital with worse-than-average outcomes claims that unmeasured patient-related factors account for the poor results, that claim may be worth more detailed investigation.

If high volumes for a particular procedure or diagnosis are primarily the result of superior outcomes, then the argument for volume data is even stronger. Since published hospital mortality data have only recently become available (see ch. 4), a relationship between volume and outcome implies that physicians (and possibly patients) have been able to use informal qualitative measures to guide more patients to physicians and hospitals with better results. Primary care physicians may consider both the mortality and other complications of their patients referred to certain specialists. Observations in the operating room or at the bedside may also alter one's confidence in the quality of care provided by specific physicians. Although such methods may be somewhat haphazard, they allow for a wide range of implicit but important criteria that may be valuable in the identification of which providers to seek out and which ones to avoid. It would be impossible to collect and make available such data, but if selective referral occurs, then the observation of a higher than expected volume of patients with diagnosis X in a hospital may be a valuable indicator of better-than-average quality.

It is important to note, however, that to use volume as an indicator of the quality of care, one must control for the various factors that influence volume. Large hospitals, for example, tend to have more patients of most diagnoses than small hospitals, irrespective of their relative quality. Public hospitals tend to treat a disproportionate share of diagnoses common among poor people. Selective contracts between certain payers and hospitals will also alter volumes. In much the same way that hospital-specific mortality rates are meaningless as outcome indicators until adjusted for case mix and certain other factors, hospital volumes are meaningless until adjusted for factors such as size of hospital, ownership, medical staff, and selective contracts. Although analyses with such adjustments have not yet been undertaken, they may be worth pursuing, especially for diagnoses and procedures for which there is evidence of selective referrals.

One additional use of volume as an indicator of the quality of care arises from the possibility of a volume-outcome relationship for physicians. Fewer studies have examined the volume-outcome relationship for physicians than have examined it for hospitals. Furthermore, the results for physicians are less consistent than those for hospitals, although some of the inconsistency may be due to methodological problems that can be overcome with better data and more analysis. Moreover, the problems of chance variation in small numbers of patients would make physician-specific data on mortality rates even less reliable than hospital-specific data. Volume data for physicians, however, may be far less controversial than outcome data. Thus, work on the volume-outcome relationship and familiarity with the use of hospital data could help set the stage for the use of physician volume data as an additional guide for consumers.

In choosing a physician or hospital, consumers should not just "go by the numbers. " Instead, if there is good evidence of a volume-outcome relationship for the patient's specific diagnosis or prospective procedure, the patient should discuss the information with a primary care physician. Suppose, for example, that a physician is recommending that a patient have CABG surgery and there are several hospitals in the community with open-heart surgery teams. Even if hospital-specific mortality data are available, there may be questions as to how they should be interpreted if none of the hospitals have significantly high or low mortality rates. As proximity is not a major issue if there are several local hospitals and if the mortality rate (3 to 5 percent) is not trivial, the patient may want to find the best, or at least avoid the worst, institution.

Suppose the hospital initially selected had a low (but not significantly so) mortality rate last year, but this rate was based on only a small number of cases. If this hospital also had a low volume, it would be reasonable to press the physician on whether one of the higher volume centers with comparable mortality rates might not be more likely to have a lower true risk of a poor outcome. Such a question may encourage the physician to think further about the referral and perhaps informally seek out additional information about the best hospital to send the patient. Although this is a rather "soft" use of information, it is probably commensurate with the precision of the available data.

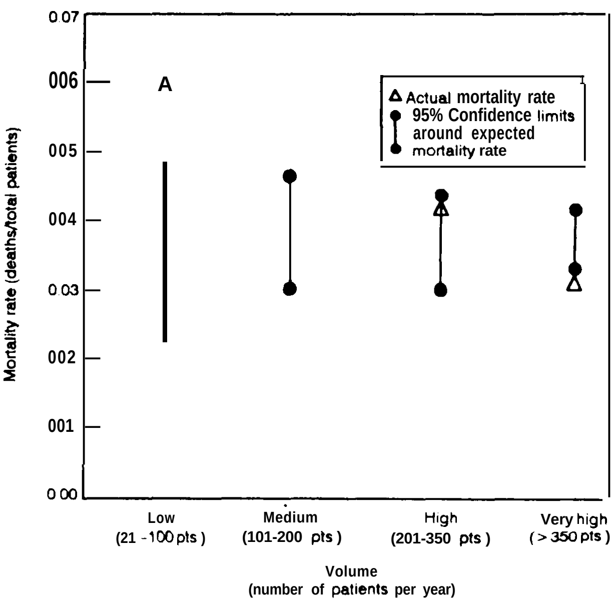
In using information about the relationship between volume and outcome, it is important to know the form of the curve for a particular diagnosis or procedure. In the analysis in this chapter, all findings with dichotomous results and with "downward-sloping, " "L-shaped," and "U-shaped" curves were grouped together. If there truly is a "U-shaped" curve, then it is necessary to identify the volume level above which mortality rates begin to worsen. Several studies have estimated "U-shaped" curves, but none have tested whether a "U" was really superior to an "L" or similar form. Nor did the studies find much evidence that very high-volume hospitals actually had worse results. The only exceptions are the studies of outcomes for newborns by Rosenblatt, et al. (538) and Williams (702). In both instances, the authors argued that the apparently worse outcomes for newborns in the very high-volume hospitals could be attributed to the very high-risk infants referred to those hospitals pursuant to perinatal regionalization policies. Unless additional studies provide clear evidence that worse outcomes occur in very high-volume centers, the public need not worry too much about reports of "U-shaped" curves.

Even if outcomes do not get worse in very high-volume hospitals, available volume-outcome studies do not necessarily imply that more is better. In many instances, the rule might be: Avoid the very low-volume setting; once you find a hospital with a volume of X, there is little to be gained by looking for a hospital with higher volume. To make recommendations about specific optimal volumes would require analyzing up-to-date data on specific diagnoses and procedures across a wide range of hospitals. Unfortunately, the available published studies do not present such analyses, but the data are generally available and it would be relatively simple for an experienced research group to undertake the necessary analyses and make public the findings.

To provide a sense of how data might be presented, consider figure 8-4. (Similar data are published in a consumers' guide in the Washington, DC area (693).) The figure indicates age- and sex-adjusted mortality rates for patients undergoing CABG surgery in hospitals with various volumes and also shows the confidence intervals, the ranges in which mortality rates would be expected to fall if volume were not a factor. (Although adjusting for risk factors other than age and sex would improve the quality of the data, the presentation could be similar.) Mortality rates in the very highest volume hospitals are significantly lower than expected; part of the reason is that at higher volumes, the confidence interval narrows. Because hospital-specific mortality data are more reliable at high volumes, however, the volume data for hospitals with high volume are less valuable. Also, patients will be less willing to switch hospitals for the relatively small incremental improvement in expected mortality associated with very high-volume, in contrast to medium- or high-volume, hospitals.

Figure 8-4 also shows that patients undergoing CABG surgery in low-volume hospitals experience significantly higher than expected mortality rates. The difference not only is statistically significant, but it amounts to a half-again higher

Figure 8-4. — Comparison of Actual and Expected Mortality Rates for Patients Undergoing Coronary Artery Bypass Graft Surgery in California, 1983



SOURCE: J A Showstack, K E Rosenfeld, D W Garnick, et al, Institute for Policy Studies, University of California, unpublished data, San Francisco, CA, 1987 rate—a 6-percent mortality rate instead of a 4-percent rate. More importantly, because of the problems of chance variability in mortality rates, review of hospital-specific mortality rates would identify few of the low-volume hospitals as having significantly poor hospital-specific outcomes. Thus, both hospital-specific mortality data and more general volume-outcome information are helpful in guiding consumers to ask better questions of their physicians.

The use of volume and outcome data varies with the specific situation at hand. In many situations, hospitalization and treatment must be immediate, and there is little time for discussion, let alone referral of a patient to other settings. In other situations, however, there may be time for reflection and discussion, but the evidence may suggest only a very weak relationship between volume and outcome. Although this relationship may be statistically significant because of the large data sets used for the analysis, the difference between an average mortality rate of 1.0 percent and 1.1 percent may not be worth pursuing for some patients, especially since there may be other factors of importance, such as proximity, the retention of a well-trusted family physician, or an institution's reputation for having attentive and responsive nursing staff.

CONCLUSIONS AND POLICY IMPLICATIONS

OTA's review of the research literature on the volume-outcome relationship for hospitals and physicians suggests that, at least for some diagnoses and procedures, higher volumes are associated with better outcomes. For 13 procedures and diagnoses reviewed in OTA's literature survey, more than half of the studies focusing on hospital volume showed this relationship. For only two procedures, femur fractures and stomach operations, did a majority of studies show no relationship between volume and outcome. The evidence for hospitals overwhelmingly showed worse outcomes at lower volumes for CABG surgery, intestinal operations, total hip replacement, cardiac catheterization, abdominal aortic aneurysm, and biliary tract surgery. Fewer studies focused on physician volume than on hospital volume, and more of the studies on physician volume either had inconsistent findings or showed no effect of volume on outcome.

To some extent, it is difficult to determine whether volume is a useful indicator of the quality of care because of the continuing controversy over the relative importance of 1) increased volume's providing the opportunity for practice and thus better outcomes, and 2) intrinsically better providers' generating increased volume through referrals. The repeated observation of a simple association between volume and outcome does not help distinguish between these two hypotheses or reveal any other causal mechanisms.

Photo credit: Cleveland Clinic foundation

Lower volume of coronary artery bypass graft surgery in hospitals was associated with higher mortality rates in 11 of 14 studies reviewed by OTA.

Regardless of the true causal pathway, volume information is useful as an indicator of quality. If the only influence is of practice, consumers would usefully be directed toward more experienced practitioners. If the influence is primarily from good outcomes' generating higher volumes, then volume may be even more valuable to consumers, because high volumes generated by selective referrals may be the best indicator of good quality.

Research focused in specific areas could provide necessary further information about the volume-outcome relationship. A problem with much of the research to date has been its academic focus; investigators have explored various analytic questions rather than developing sets of estimates that are directly useful for consumers. For example, although many studies indicate the presence of a volume-outcome relationship, the wide range of analytic methods and variable specifications makes it difficult to determine whether poor outcomes are concentrated at very low volumes or whether improved outcomes are seen throughout the observed range of volumes. Further studies are required to determine whether the recommendations should be to "seek the highest volume center" or to "avoid places with fewer than X patients. "

To some extent, the variety of functional forms and approaches used by various investigators reflects the constraints of the available data. A very useful study would compare the findings of studies that used the same analytic techniques on various types of data for patients who had the same diagnosis or procedure. For example, one can obtain data on post-discharge mortality and readmission for Medicare patients, but using these data limits the analysis to patients over age 65. Are linked inpatient and ambulatory data superior to data on inpatient outcomes? In a similar vein, do hospitals with high rates of other complications also have high mortality rates? Do these objective measures match other evaluations of quality, such as those developed by the peer review organizations?

The quality of the data is probably more important for the evaluation of specific hospitals than for the analysis of volume-outcome issues. Additional data that may improve the certainty of a judgment with respect to quality of a particular hospital are very important because of the potentially disastrous consequences of misclassification. In contrast, random noise in the data used for volume-outcome studies merely makes it somewhat more difficult to detect what is going on; a larger sample size can often overcome the problem.

The evidence of a relationship between physician volumes and outcomes is less clear than the evidence for hospital volumes and outcomes, and none of the existing studies of physician volumes is fully convincing. Prior research has been constrained by both data and methodological problems. Some newly available data sets are now including physicians' license numbers (and in Arizona's case, physicians' names), so it will be possible to identify a physician's patients across several hospitals.

Another crucial question that remains to be resolved is whether high volumes arise from selective referrals of patients to hospitals and physicians with better-than-average outcomes, whether better outcomes arise from high volumes, or whether both phenomena arise in some complex relationship. Methodologically, this question is a difficult one to address, but various simultaneous-equation techniques and better data on patient referrals may provide more convincing evidence.

Both selective-referral and practice-makes-perfect effects have a time dimension. The fact that a beginning surgeon will eventually perform 200 procedures during the first year of practice may not affect the outcomes of the first patients on which he or she operates in that year. It is often assumed, however, that a hospital or physician with a volume of 200 procedures in a year had about that many procedures in prior years. The implicit assumption is that all volumes have reached some steady-state level. In reality, new physicians enter practice, new procedures are developed, hospitals offer new services, and past volume levels may differ from current (and future) ones. What is the shape of the personal and institutional learning curves after a new procedure or treatment is introduced? What yearly volume is necessary to keep skills from deteriorating?

84-752 o - 88 -- 7

For considering the selective referral hypothesis, timing is also important. If outcomes or reputations influence referrals, what is the time lag involved? Is an occasionally higher-than-expected mortality rate ignored, while only consistently better or worse than average results affect referrals? Can a hospital that replaces a poor-quality surgeon with a good-quality one increase its volume, or is a poor reputation difficult to erase? Likewise, for how long can a hospital (or physician) with deteriorating outcomes maintain old referral sources? These questions have not been explored in any empirical studies to date.

Finally, a series of very detailed studies could explore precisely what clinical factors account for differences in outcomes, in effect, to validate the observed relationship between volume and outcome. Such studies would probably rely on careful review of patients' charts from various settings to determine the relative importance of errors of commission and omission, differences in technique, monitoring, support, and the like. It is probable that the importance of various factors will depend on the procedure or diagnosis studied.

Even with the substantial gaps in knowledge about the volume-outcome relationship, there are still policy measures worthy of consideration. In discussing various policy options, it is important to consider unintended incentive effects. The following five policy measures are ordered roughly in terms of increasing strength of incentives for—and ability of—hospitals to manipulate the data or otherwise behave in undesirable ways.

Educating the general public about the relationship of lower hospital volumes to worse outcomes is the simplest approach. Even if the causal linkages are not clear, it seems reasonable to argue that, in the absence of other evidence, hospitals with high volumes are preferable to nearby ones with very low volumes. Upon receiving a referral for a specialized procedure, an informed consumer might then ask his or her primary care physician about the volume and quality of the proposed specialist and hospital, given the relevant alternatives. Educating the general public would impose no new data collection requirements and the potential costs are small. One could easily see an educational strategy implemented through articles in the lay press, such as ***Readers' Digest*** or the Sunday newspaper supplements

A second level of intervention might be directed toward physicians through their specialty associations and continuing-education programs. [6] Specialty associations might be encouraged by Congress to collect volume and outcome information in their areas and make it available to physicians. In particular, these associations could focus on some of the more individualistic and sensitive factors that may improve physicians' ability to selectively refer patients to settings and physicians with better outcomes. It might be necessary to clarify whether such educational efforts by local specialty associations would raise antitrust problems.

A third level of intervention would be for States or other State-level entities to require the routine collection and publication of hospital-specific volume information. For the 28 States with mandatory hospital discharge abstract reporting requirements, this task would be an easy one. States could clearly not publish data for all hospitals and all procedures and diagnoses, but selected data could be made available to interested parties. Selected hospital-specific information could be reprinted by local newspapers. California Blue Shield published a list of hospitals with their CABG surgery volumes (112), and Blue Cross and Blue Shield of Ohio published a Consumer Guide with the number of patients by DRG (73). Some consumer organizations and magazines have done the same (693).

Requiring the disclosure of hospital-specific volumes is a measure that must be carefully considered because of potential unanticipated effects. Consider, as an example, a fourth level of intervention whereby a hospital is penalized financially by third-party payers, or a particular unit shut down by regulators, if certain volume levels are not maintained. This approach would create incentives for hospital administrators to make sure that at least the minimum acceptable number of patients are treated. One could imagine memos

[5] Information-dissemination strategies are discussed in ch. 2.
[6] Physician specialty boards are listed in ch. 10.

from hospital administrators to the medical staff pointing out that if another 20 patients are not operated on before the end of the fiscal year, the X unit will be closed down. This pressure might lead to the relaxation of standards for the appropriateness of an admission. Moreover, basing payment or regulatory decisions, which affect a hospital's ability to continue in a specific line of business, on volume may not be fair because volume is at best merely a proxy for quality.

A fifth policy application using volume data as an indicator of the quality of care is in the realm of selective contracting. Insurers, health maintenance organizations, and other agents such as Medicaid programs may wish to steer the patients for whom they are responsible to hospitals that are likely to achieve better outcomes. If reliable outcome data are available either through sources that routinely collect data or through carefully structured bids, then for high-volume hospitals, outcome data may be preferred to simple volume data because the outcome data would include only a small chance component. For low-volume hospitals, the outcome data tend to be too unreliable. On the other hand, if outcome data are unavailable or too subject to manipulation, then volume of specific procedures may be a proxy for quality. (For example, suppose an agency were to announce that it was going to utilize hospital discharge abstracts to determine death rates for the purposes of contracting. A hospital with a high inpatient mortality rate may monitor patients for complications and transfer those at risk of death, thereby improving its own statistics. It would be far more difficult to manipulate volume figures, and it is unlikely that many hospitals could attempt such a strategy without detection.)

Additional policy applications depend on a better understanding of the relationship between volume and outcome. For example, if increasing volume for specific procedures or diagnoses does lead to improved outcomes, then the argument for explicit regionalization strategies becomes far stronger. If hospital volume is far more important than physician volume, then one would argue against the peripatetic surgeon. On the other hand, if physician volume is the crucial variable, then "circuit riding" may become far more common, with many low-volume hospitals sharing a single high-volume physician. If higher hospital malpractice claims are associated with lower volumes, then malpractice insurance premiums should be adjusted to reflect this risk factor. All of these and other options must await future research. Fortunately, many of the policies directed toward consumers do not require much additional information.