Interpreting the Volume-Outcome Relationship in the Context of Health Care Quality: Workshop Summary

by Maria Hewitt for the Committee on Quality of Health Care in America and the National Cancer Policy Board

ISBN: 0-309-55831-X, 128 pages, 8 1/2 x 11, (2000)

Interpreting the Volume–Outcome Relationship in the Context of Health Care Quality

Workshop Summary

by Maria Hewitt
for the
Committee on Quality of Health Care in America
and the
National Cancer Policy Board

INSTITUTE OF MEDICINE
Washington, D.C.

INSTITUTE OF MEDICINE • 2101 Constitution Avenue, N.W. • Washington, DC 20418

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

Support for this project was provided by The Agency for Healthcare Research and Quality.

Additional copies of this report are available in limited quantities from the National Cancer Policy Board, 2101 Constitution Avenue, N.W., Washington, DC 20418. The full text of this report is available at **www.nap.edu.**

For more information about the Institute of Medicine, visit the IOM home page at **www.iom.edu.**

Printed in the United States of America.

The serpent has been a symbol of long life, healing, and knowledge among almost all cultures and religions since the beginning of recorded history. The serpent adopted as a logotype by the Institute of Medicine is a relief carving from ancient Greece, now held by the Staatliche Museen in Berlin.

"Knowing is not enough; we must apply.
Willing is not enough; we must do."
—Goethe

INSTITUTE OF MEDICINE

Shaping the Future for Health

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. William A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Kenneth I. Shine is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. William A. Wulf are chairman and vice chairman, respectively, of the National Research Council.

REVIEWERS

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the NRC's Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. We wish to thank the following individuals for their review of this report:

Bruce E. Bradley, General Motors Corp., Detroit, Michigan
John S. Chipman, Ph.D., Department of Economics, University of Minnesota
Judith H. Hibbard, Dr.P.H., Department of Planning, Public Policy, and Management, University of Oregon
David Nerenz, Ph.D., Director, Health Care Studies, Institute for Managed Care, Michigan State University
Donald Nielsen, M.D., American Hospital Association, Chicago
Kathryn A. Phillips, Ph.D., Associate Professor of Health Economics, School of Pharmacy, Institute for Health Policy Studies, University of California at San Francisco

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations nor did they see the final draft of the report before its release. The review of this report was overseen by Sheldon Greenfield, Ph.D., Director, The Primary Care Outcomes Research Institute, New England Medical Center Hospitals, Boston, appointed by the Institute of Medicine, who was responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report rests entirely with the authoring committee and the institution.

Contents

Interpreting the Volume–Outcome Relationship in the Context of Health Care Quality: Workshop Summary

ABSTRACT

The Institute of Medicine (IOM) held a workshop on May 11, 2000, to review the current understanding of the relationship between volume of health services and health-related outcomes; assess the potential health policy implications of the use of volume as an indicator of quality of care; and identify needed areas for future research relative to volume and quality of care. The workshop was jointly sponsored by IOM's Committee on Quality of Health Care in America and the National Cancer Policy Board, with financial support from the Agency for Healthcare Research and Quality. The workshop was structured around presentations of two background papers: a cross-disciplinary, systematic review of the published research literature conducted by Drs. Ethan Halm, Clara Lee, and Mark Chassin of the Mount Sinai School of Medicine and a synthesis of the literature relevant to policy issues by Dr. Adams Dudley and colleagues from the University of California, San Francisco.

A higher-volume, better-outcome association was observed in three-quarters of the studies reviewed. Volume is, however, an imprecise indicator of quality. Some low-volume providers have excellent outcomes, and conversely, some high-volume providers have poor outcomes. Volume per se does not lead to good outcomes in health care; it is instead a proxy measure for other factors that affect care. With few exceptions, however, the literature does not shed light on the structures or processes of care that underlie the apparent relationship.

Many workshop participants noted that volume may be the best available proxy indicator of quality for certain conditions, but efforts should be made to accompany volume information with other quality indicators and with explanatory information. Furthermore, several participants suggested that when research confirms a volume–outcome link, this information should be disclosed to the public to support health care decision-making. In making such disclosure, however, the limitations of the data and how to interpret the information must be made clear to the intended audience. Public release of data on volume may motivate providers to report process and outcome data, thereby laying the ground for the next generation of health care quality measures. When better measures of quality of care than volume are developed, they should replace the volume measures.

A wide-ranging set of research topics—from policy research and demonstration programs, to basic methodological research—was outlined by workshop participants to elucidate why the relationship between volume and outcome exists, and how best to implement policies to improve health care.

INTRODUCTION

An association between higher volumes and better outcomes has been well documented for certain types of health care. Some health care purchasers are using these research findings to refer patients to higher-volume settings for selected procedures. The evidence prompted the Na-

tional Cancer Policy Board to recommend that cancer patients in need of highly complex surgical procedures go to higher-volume facilities for care (IOM, 1999). Despite considerable evidence of a volume–outcome relationship, many questions remain about the nature of the relationship, the processes of care that might explain it, and its implications for health care policies.

On May 11, 2000, the Institute of Medicine (IOM) held a workshop[1] to bring together experts to:

1. review evidence of the relationship between volume of services and health-related outcomes;
2. discuss methodological issues related to the interpretation of the association between volume and outcome;
3. assess the applicability of volume as an indicator of quality of care; and
4. identify research needed to better understand the volume–outcome relationship and its application to quality improvement.

The workshop was jointly sponsored by IOM's Committee on Quality of Health Care in America and the National Cancer Policy Board (membership listings can be found in Appendix B). The IOM received financial support for the workshop from the Agency for Healthcare Research and Quality (AHRQ). The workshop summary will inform the deliberations of its sponsoring groups. The National Cancer Policy Board will, for example, issue a White Paper including recommendations regarding volume-based quality indicators relevant to cancer care. The committee charged to develop a national quality report on health care delivery will review the summary as it assesses quality measurement systems.

The workshop was organized around presentations of two background papers commissioned for the workshop (Appendixes C and D):[2]

- "How Is Volume Related to Quality in Health Care? A Systematic Review of the Research Literature?" by Ethan A. Halm, Clara Lee, and Mark R. Chassin
- "When and How Should Purchasers Seek to Selectively Refer Patients to High-Quality Hospitals?" by R. Adams Dudley, Richard Y. Bae, Kirsten L. Johansen, and Arnold Milstein

This report summarizes the workshop presentations, discussions during the workshop, and also the main points made in the two background papers.

[1]The workshop agenda and participant list are reproduced in Appendix A.

[2]A planing group made up of members of the Committee on Quality of Health Care in America and the National Cancer Policy Board oversaw the development of the agenda and outlines for the two commissioned background papers. Members included Don Berwick, Molly Joel Coye, Mark Chassin, Diana Petitti, and Jane Sisk. The planning group's role was limited to planning the workshop. IOM staff prepared this factual summary of what occurred at the workshop.

WHY DOES THE VOLUME–OUTCOME RELATIONSHIP MATTER? WHO CARES AND HOW COULD IT BE USED?

To set the context for the discussion, Dr. John Eisenberg, in opening remarks to workshop participants, identified a diverse set of potential users of information on the volume–outcome relationship.

Health Insurance Purchasers and Health Plans

Health insurance purchasers could use findings from research on the volume–outcome relationship to stipulate "evidence-based referrals" in contracts with health plans. Such a program is being explored by the Leapfrog Group, a coalition of employers interested in improving the quality of care for employees. In addition, employers could make information about the relationship between volume and outcome directly available to employees (e.g., on a company intranet site) and encourage employees to choose hospitals and providers based on available evidence. Similarly, health plans could direct members to high-volume providers.

Consumers

Information about the relationship between volume and outcome could be provided more broadly through public websites, via advocacy groups, or as part of widely distributed quality report cards. In New York State, for example, information on the volume of cardiovascular and other procedures performed by individual surgeons and by hospitals is available through the Center for Medical Consumers, a nonprofit advocacy organization (www.medicalconsumers.org).

Insurers

Insurers must decide which new technologies or services to include in their benefit packages and, sometimes, the conditions under which these new service will be provided. For a new technology, insurers could offer interim service coverage until any volume–outcome relationship can be established. Once a relationship became apparent, coverage could be conditional on its provision in high-volume settings.

Hospital Administrators

Hospital administrators could, based on a positive volume–outcome relationship, decide to credential providers according to their volume of procedures or number of hospital admissions. Staffing decisions could also be affected. For some services, patients might be triaged to specialty units within hospitals, in part to ensure that providers have a sufficient volume to maintain their skills. One of the tenets of disease management is that assigning patients to certain disease-specific programs or institutions will improve care. This is not entirely related to volume, but evidence of a volume–outcome relationship for certain services could foster the development of specialty services, concentrating patients into settings with high volume.

Regulators

A few states have certificate of need (CON) legislation, a regulatory mechanism for review and approval of capital expenditures and service capacity expansions by health care facilities. In

New York, for example, facilities must be granted certificates of need before they can offer certain cardiovascular procedures. In part, certificates are granted on the basis of volume. States regionalize some types of care, when evidence shows that such organization improves outcomes. Trauma and neonatal intensive care are, for example, regionalized to ensure that complex cases are cared for in institutions with the facilities and personnel to manage them. Regulators could also require that volume data be reported and, in turn, made available to consumers.

Accrediting Organizations

Organizations such as the Joint Commission on Accreditation of Healthcare Organizations could consider volume in accrediting hospitals and other health care institutions. The National Committee for Quality Assurance could evaluate health plans based on their application of volume information (e.g., to inform enrollees, move care to higher-volume providers).

Professional Organizations

Physician specialty boards could certify providers, taking into consideration their experience with a certain procedure. The American Board of Internal Medicine, for example, considers experience and volume when certifying internists to provide electrophysiology services.

Medical Educators

Knowing how much experience (or volume) is needed to develop expertise could guide curriculum development and the appropriate allocation of training time for certain procedures. In the context of continuing education, research on the volume–outcome relationship could provide information on how to sustain skills or acquire new ones. Methods to train providers without putting patients at risk are needed. In surgery, animals have often been used to help providers acquire new skills, but it is likely that computer-based simulation models commonly used in the airline industry to train pilots will be forthcoming in medicine. Dr. Eisenberg noted that research will be needed to ensure that simulated volume is related to favorable outcomes.

HOW IS VOLUME RELATED TO QUALITY IN HEALTH CARE?

To assess how volume is related to quality in health care, Drs. Ethan Halm, Clara Lee, and Mark Chassin, of the Mount Sinai School of Medicine, conducted a systematic review of the published research literature (Appendix C). As part of their cross-disciplinary review, they examined evidence from 88 studies concerning eight conditions and procedures:[3]

- coronary artery bypass graft (CABG) surgery,
- pediatric cardiac surgery,
- carotid endarterectomy,
- abdominal aortic aneurysm repair,

[3]Inclusion criteria for the review included publication since 1980, analysis of a population-based sample, volume as an independent variable, and inclusion of a health outcome as the dependent variable. Excluded were studies of care that has been regionalized or is highly regulated (i.e., trauma care, newborn intensive care, organ transplantation) and conditions or procedures for which there were very few articles (e.g., knee replacement surgery).

- cancer surgery,
- percutaneous transluminal coronary angioplasty (PTCA),
- acute myocardial infarction, and
- acquired immunodeficiency syndrome (AIDS).

Their review focused on the nature of the relationship between volume and outcome, the causal pathways that might explain the relationship, and whether the research findings might apply to other clinical areas. Historically, the volume–outcome relationship has been explained by "practice makes perfect" or patient selection (i.e., "busy doctors see healthier patients"). The authors propose a conceptual model that identifies a series of factors that might underlie the relationship between volume and quality of care (Figure 1).

The authors conclude that for a wide variety of surgical procedures and medical conditions, higher volume (whether assessed by hospital or by physician) is associated with better health outcomes. The uniformity with which the published research documents or confirms the existence of this association is compelling (although the review authors caution that a publication bias against negative findings cannot be ruled out):

- Statistically significant associations between higher volume and better outcomes were found in 79 percent of the studies of hospital volume and 77 percent of the studies of physician volume.
- None of the studies reviewed showed a negative effect of volume.
- A significant volume effect was observed in all 16 of the studies judged to have the soundest research methods.

A higher-volume, better-outcome association was observed in three-quarters of the studies

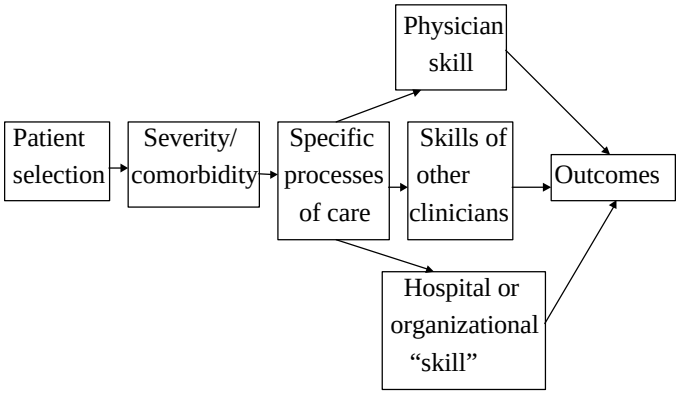


FIGURE 1 Conceptual Framework: How Could Volume Affect Quality?

reviewed. There is, however, great variation in outcomes, especially for low-volume providers. Some low-volume providers have excellent outcomes, and conversely, some high-volume providers have very poor outcomes. Furthermore, the performance gap between low- and high-volume hospitals appears to narrow with time as procedures become well established (e.g., the

volume of some cardiovascular procedures has tripled in less than a decade). Consequently, volume may be most pertinent when a new technology is beginning to diffuse into general practice.

Volume per se does not lead to good outcomes in health care. It is a proxy measure for other factors that affect care. In one of the studies, about one-third of the survival benefit among patients with acute myocardial infarction that was attributed to high-volume hospitals could be explained by the use of appropriate processes of care, for example, use of beta-blockers, aspirin, thrombolytics, and revascularization procedures. Unfortunately, most of the studies reviewed did not include information on processes or systems of care that might explain the underlying reasons for the volume–outcome relationship.

It is difficult to summarize the findings by procedure or condition because of study differences in the definition of low- and high-volume, in analytic techniques, and in adjustment methodologies. Even so, the volume–outcome relationship appears to be particularly dramatic for certain low-frequency, high-risk, surgical procedures such as surgery for cancer of the pancreas and esophagus. For these procedures, rates of short-term mortality are 2 to 3 times greater in low- versus high-volume hospitals. For other procedures or conditions under review, the volume effect was not as great or as consistent (see Appendix C for procedure- and condition-specific findings). Statistically significant differences in outcomes by volume may not always be of sufficient magnitude to be clinically significant; furthermore, other considerations can mitigate the importance of such differences. For example, if the difference between high- and low-volume hospitals' surgical mortality rates was only 1 or 2 percentage points, and the underlying condition for which the surgery was performed had a poor long-term prognosis, it may not be prudent to implement policy based only on a statistically significant difference in short-term mortality.

The authors caution that they may not have identified all relevant studies of the relationship between volume and outcome and that many studies, particularly those involving cardiovascular procedures, were conducted in New York and thus may not generalize to states with less regulation of such procedures. Negative publication bias (researchers' failing to report the absence of a relationship between volume and outcome or journals' failing to publish negative results) could also overestimate the positive nature of the volume–outcome relationship.

Drs. Halm and colleagues judged the methodological strength of the research using an explicit quality scoring method[4] and concluded that the methodological rigor of the studies was modest, limiting generalizability and leaving many unanswered questions about the nature and causes of the observed association:

- It is unclear whether volume thresholds exist above which outcomes are better. For example, is it only the busiest providers that have the best outcomes, or do all but the lowest-volume providers have better outcomes?
- It is likely that the effects of physician and hospital volume combine or interact. One study suggests a multiplicative effect leading to much better outcomes among high-volume physicians practicing in high-volume hospitals. The relative contributions of physician and hospital volume to outcomes, however, have been examined in only a few studies.
- Studies generally do not illuminate *how* experience with procedures (or conditions) that are closely related to the procedure under study affects outcomes. Are cognitive and technical

[4]The scoring method considered the representativeness of the study sample, sample size, number of adverse events, unit of analysis (e.g., joint effects of physician and hospital volume), inclusion of clinical processes of care, appropriateness of patient selection, categorization of volume, quality of risk adjustment, and outcomes measurement.

skills, for example, transferable from one cardiovascular procedure to another (i.e., what should "count" toward volume?)?

- Longitudinal studies are needed to answer some questions. Once high volume is attained, for example, does it have to be sustained, or can lower volumes be adequate to maintain skills?
- The literature focuses on the effects of the attending physician's care. The experience of collaborating physicians and other clinical personnel has been far less thoroughly examined (e.g., do members of a surgical team need to achieve a team volume?).
- With few exceptions, the available literature does not identify clinical processes of care that contribute to observed differences in outcomes.
- Most studies focus on short-term outcomes such as in-hospital mortality. No studies examine longer-term outcomes such as quality of life or functional status.

Five workshop participants were invited to provide reactions to the background paper:

1. Norman Hertzer, a vascular surgeon from the Cleveland Clinic;
2. Edward Hannan, a health services researcher at the State University of New York at Albany;
3. Colin Begg, an epidemiologist from the Memorial Sloan-Kettering Cancer Center;
4. Bruce Hillner, an internist and health services researcher at Virginia Commonwealth University; and
5. Arnold Milstein, medical director at the Pacific Business Group on Health and chief physician at William M. Mercer, Inc.

Dr. Hertzer pointed out that the difference between low and high volumes for vascular surgery is sometimes small (e.g., fewer than 3 versus more than 12 procedures for carotid endarterectomy; fewer than 4 versus more than 10 procedures for aortic aneurysm repair). While recognizing that the outcomes of low-volume providers are significantly worse than those of high-volume providers, he noted that providers in the intermediate-volume group often are indistinguishable from either low- or high-volume providers. It is often difficult to isolate the factors responsible for the volume–outcome relationship because, in the case of vascular surgery, different types of surgeons may be involved (e.g., general surgeons, neurosurgeons). Furthermore, the same surgeon may be responsible for a high volume of cases in a large, high-volume hospital and for the cases performed in a small, low-volume hospital. These observations underscore the importance of simultaneously tracking both physician and hospital volume and controlling for type of surgeon. The difficulty of defining surgical experience with volume further complicates these analyses. Surgeons with low volume can include those who are newly trained, but may have state-of-the-art skills, as well as surgeons with years of experience and good outcomes who are reducing their work load before retirement.

The quality of available data systems is central to monitoring the quality of health care, and Dr. Hannan described a hierarchy of data for quality research. Optimally, one would be able to track the use of processes of care known through well-designed studies (e.g., clinical trials) to affect care. When accurate data are available to show that a certain procedure is, for example, being under- or overused, it is relatively easy for quality improvement programs to inform providers and hold them accountable. For most clinical care, however, strong links between processes and outcomes are not well established, and even when they are, there are relatively few data systems in place to track adherence to evidence-based practices.

In the absence of good data on processes of care, data about outcomes that include risk adjustment using detailed clinical data (usually available only in the medical chart or specialized databases) provide the best measurement of quality of care. In New York and New Jersey (and soon in California), statewide clinical databases are available for cardiac surgery, allowing analyses of outcomes by both individual surgeons and hospitals. Risk-adjustment methodologies, however, have not yet been perfected, so the interpretation of findings can be difficult, especially when applied to individual providers.

Next in the hierarchy of quality-of-care measures are outcomes for which only administrative data (e.g., hospital discharge databases) are available to perform the risk adjustment. Such measures may suffer from both the specificity of the assessment of severity of illness (in general, risk-adjustment models using administrative data will have less predictive power than models based on detailed clinical data) and from reporting problems that plague administrative databases. While outcomes data can be informative, they are often limited to short-term measures such as death or complications, and other, longer-term outcomes of interest to patients are usually not available (e.g., functional status, quality of life).

Dr. Hannan suggested that data on volume can serve as a proxy for quality but that its value as a quality indicator can be enhanced by using it with other imperfect indicators, such as adjusted or unadjusted mortality data from administrative databases.

It is often difficult to judge when to implement policies based on research findings. Dr. Begg identified four criteria to consider in assessing the strength of the evidence on the volume–outcome relationship and adopting it as a standard of quality of care:

1. The relationship must be plausible and logical.
2. The observed trend must be consistent.
3. The size of the outcome difference must be substantial and meet stringent statistical criteria.
4. The effect must be confirmed in multiple studies.

In his view, these criteria are met for certain procedures included in the literature review (i.e., surgery for cancer of the pancreas and esophagus).

Dr. Begg agreed with other workshop participants when they encouraged the examination of processes of care that might explain the volume–outcome relationship, but cautioned that it may not always be possible with available data to identify the relevant processes, especially for procedures that exhibit small volume–outcome trends. In the case of cancer, the administration of adjuvant therapy may be documented in administrative databases, but not the dosage and whether the course of therapy was completed. Dr. Begg conducted a study of the relationship between the volume of surgery for colon cancer and outcomes, showing an outcome difference for stage III colon cancer. Investigators hypothesized that the trend would be explained by better adherence to recommendations on adjuvant chemotherapy use by higher-volume hospitals. When data on adjuvant therapy were retrieved and analyzed, however, appropriate use of adjuvant therapy did not explain the association. Any number of unmeasured or unrecorded processes of care could account for these findings. Surgical judgement, skill, improvisation in the face of unpredictable events, and the duration of surgery may, for example, all contribute to the finding, but they are difficult to study. Dr. Begg also described the difficulties of accurately assessing patient selection factors with available data (e.g., frailty is not necessarily captured in administrative data beyond lists of comorbid conditions).

Dr. Hillner presented data to show that any policies directing patients to high-volume centers would affect many people, because lower-volume facilities account for a significant fraction of patient care. In Maryland from 1992 to 1996, for example, 36 percent of surgeries for colon cancer were performed by surgeons treating fewer than 5 cases per year (average, 2 per year). Three-quarters of the surgeons performed 10 or fewer surgeries per year. By contrast, in Ireland where care is concentrated in high-volume settings, 80 percent of surgeons who do colectomies perform 10 or more surgeries for colon cancer per year. Dr. Hillner suggested that most individuals, when informed of the relationship between volume and outcome, would not find current practices in the United States acceptable.

Dr. Hillner suggested some parallels between programs to promote selective referral for surgical procedures and disease management programs for patients with chronic conditions. For both, patient care resources are concentrated. Establishing the relationship between volume and outcomes for hospital-based procedures has been somewhat easier because of the availability of large databases containing information on hospital outcomes. There is somewhat less evidence of the success of disease management because of the limitations of available data and the need to examine longer-term outcomes of chronic diseases such as diabetes.

Dr. Milstein observed that causal models to explain favorable hospital outcomes are in a primitive state. He supported the notion that process measures of quality would be preferable to volume and the consequent need for a significant expansion of routinely collected data elements. Dr. Milstein noted, however, major resistance to such expansion on both national and state levels (especially data to assess the performance of individual physicians). Dr. Milstein questioned the assumption that the most valuable use of volume–outcome research will be for quality improvement. This assumption reflects a faith that the health industry's weak quality improvement capabilities will improve and that little of the association of volume with outcome will ultimately be rooted in volume-influenced proficiency. It may be the case that the frequency of real-world exercise of physical and mental skills will remain important, even after low-volume hospitals are made aware of the "tricks" of the high-volume trade. Since hospital performance differences are likely to persist after provider-to-provider knowledge transfer, Dr. Milstein recommended that social science research be expanded to improve understanding of consumers' use of quality indices in choosing hospitals and physicians. He noted that the existing literature suggests that the use of provider quality comparisons by consumers is not likely to result from today's "information dump of statistical tables." This will require simplification of information, putting it into multiple formats, and having it conveyed by trusted sources. Lastly, although Dr. Milstein agreed with the need for caution in using volume-based indicators, he questioned whether any workshop participants would commit to personally receiving complex treatments at the nearest low-volume hospital in their health plan's network. He suggested that this is what we currently tolerate for less well educated Americans who live near low-volume hospitals.

Discussion moderator Dr. Sisk asked workshop participants to consider the types of measures needed to implement policy and to assess the adequacy of risk-adjustment methodologies. Two main applications of studies of the volume outcome relationship were identified in the discussion. The first is research, identifying whether a relationship exists and the factors that affect the relationship. For these purposes, available risk-adjustment methods are adequate, even when applied to administrative data. According to Dr. Halm, the literature review suggested that risk adjustment, although important, did not greatly affect the magnitude of the association between volume and outcome. The second application is to provide information to individual providers of care as part of a quality improvement program or to inform consumers of the volume or out-

comes of providers in their community. Here, Dr. Hannan noted that the critical issue is the source of data, not the method of risk adjustment. Audited clinical data are needed, and administrative data are inadequate. Dr. Chassin suggested that an important consideration for providers is inclusion in the clinical database of risk factors that they view as important to outcomes. Without these components, providers will not have confidence in the performance reports.

Optimally, an information system used as part of a quality improvement program would include data on both processes of care and outcomes so that providers would know what sort of corrective actions were needed to improve outcomes. Process measures are usually lacking, however, and providers receive performance data in the form of risk-adjusted outcomes. Providers thus lack a roadmap for what to do. Dr. Chassin reported that experience in New York suggests that low-ranking hospitals have used very different strategies to improve performance (e.g., changes in personnel, processes, organization). In some regions, specialty providers have conferred with one another using a "reverse-engineering" approach to help identify differences in performance that account for outcome differences (e.g., the New England cardiovascular project).

Dr. Berwick pointed out that if quality problems stem from a process error, the steps necessary to correct the problem are relatively straightforward. If, however, poor quality stems from poor underlying skills, it may be difficult to enhance a provider's skills, although professional development training or a mentoring process could be used. Mentoring approaches have been effective in the adoption of new technology, where providers mentored by experienced practitioners helped novices move up the "learning curve." Simulation models would also be helpful, although relatively few examples of this approach are available.

In concluding the morning session, Dr. Chassin observed that evidence suggests that poor performers could do much better, but it is unlikely that all providers will ever achieve the highest level of performance.

WHAT ARE THE POLICY IMPLICATIONS OF A VOLUME–OUTCOME RELATIONSHIP?

Dr. Dudley and colleagues from the University of California, San Francisco, in their synthesis of the literature relevant to policy issues (Appendix D), put the volume–outcome relationship and its implications into the broader context of quality improvement. Dr. Dudley outlined the two general approaches to quality improvement:

1. Have providers improve their quality (including correcting overuse, underuse, and misuse).
2. Have patients switch to higher-quality providers.

The authors listed a number of specific strategies to improve quality of care that could incorporate considerations of volume:

- regulatory strategies (e.g., CON, regionalization);
- improving the skills of health professionals (e.g., physician education); and
- competition-based strategies: consumer-oriented approaches (e.g., report cards) and purchaser initiatives (e.g., selective referral, quality bonuses).

The competition-based strategies depend upon public disclosure of data to consumers and/or groups that purchase health insurance (e.g., employers). Given that the volume–outcome

relationship holds on average, Dr. Dudley pointed out that most individuals could benefit from this information. On a case-by-case basis, however, the information will not have its intended consequences for a significant share of patients. Dr. Hannan provided two examples of the risk of making referrals purely on the basis of volume. He noted that in New York, one-third of high-volume surgeons (i.e., those performing at least 150 procedures per year, or 450 procedures over 3 years)[5] had higher-than-average risk-adjusted mortality rates. Also among the 18 hospitals performing very high volumes of angioplasty (i.e., at least 400 procedures per year),[6] 8 hospitals had above average risk-adjusted mortality rates. As these examples illustrate, a relatively large share of patients would be cared for by providers with inferior performance if referrals were made solely on the basis of volume.

Dr. Dudley noted that providing consumers with information to support decision-making has been somewhat disappointing. Report cards can be difficult for consumers to understand and, where they have been published, appear to have been used by only a small percentage of consumers. New formats and communication tools are needed to improve their use, and Dr. Dudley suggested that a *Consumer Reports* model be tested for disseminating information. *Consumer Reports* represents a trusted, unbiased source of information that provides a wealth of detailed information, as well as easy-to-use summary measures with which to judge products. According to Dr. Dudley, the inclusion of information on functional status, long-term outcomes, and common chronic conditions (e.g., those treated in outpatient settings) would likely increase the relevance of such reports to consumers.

Large purchasers are beginning to hold systems of care accountable for quality improvement. The Pacific Business Group on Health (PBGH), a large purchasing coalition, provides condition-specific volume data for area hospitals on its website (www.healthscope.org) along with guidance on how to interpret the data. PBGH also requires health plans to ensure that patients with selected conditions go to high-volume hospitals. This approach is also being taken by the Leapfrog Group, a newly formed organization comprising many large employers and purchasing coalitions (including PBGH). The Leapfrog Group has developed a set of health plan performance standards that include volume standards for specific conditions, but it will recommend better measures when they become available

Information about volume can be applied without public disclosure, for example, within systems of care for quality improvement programs. Low-volume providers may withdraw voluntarily to avoid scrutiny or be motivated to examine and improve internal structures or processes of care.

Dr. Dudley and colleagues described several potential barriers to implementing a selective referral program based on a volume standard:

1. a potential for decrements in quality at higher volumes;
2. patients' preferences for care close to home;
3. patients' lack of resources to travel to hospitals that are far away;
4. patients who need immediate treatment or are too unstable to transfer;
5. loss of access in areas where low-volume services have been closed (e.g., cardiac surgery);
6. resistance from physicians and hospitals to cooperate in quality monitoring efforts; and

[5]According to these criteria, 30 percent of cardiac surgeons in New York are high-volume providers.

[6]The American College of Cardiology recommends that hospitals perform at least 200 angioplasty procedures per year.

7. effects on marketplace structure and competition:

- increased market power of high-volume hospitals (e.g., prices could rise),
- barriers to entry of new competitors (e.g., it is difficult to start at high volume), and
- potential for medically inappropriate admissions to boost volumes to meet cutoffs.

The following discussants responded to Dr. Dudley's remarks:

- Consumer perspectives—Ellen Stovall, National Coalition of Cancer Survivorship and National Cancer Policy Board; and Art Levin, Center for Medical Consumers
- Purchaser perspectives—Bruce Bradley, General Motors; and Stephen Clauser, Health Care Financing Administration
- Managed care perspective—George Isham, HealthPartners; and Sam Ho, Pacificare Health Systems
- Provider perspectives—Don Nielsen, American Hospital Association
- Research perspectives—Irene Fraser, Agency for Healthcare Research and Quality

Consumer Perspectives

Ms. Stovall, as a cancer care consumer, raised concerns about institutions' advertising on the basis of volume without other quality data or explanations of the meaning of volume and without descriptions of how the information should be interpreted.

Mr. Levin described how the New York State Department of Health has taken risk-adjusted mortality data on cardiovascular procedures in New York to produce physician- and hospital-specific reports for consumers. Mr. Levin questioned the relevance of these data to contemporary practice given that as of May 2000, mortality data were limited to the years 1994 to 1997. Volume data provided by the Center for Medical Consumers are more current—1998 hospital- and physician-specific volume data were posted on his organization's website (www.medicalconsumers.org). Mr. Levin noted that efforts to disseminate quality data to the public in New York have met with fierce resistance from physicians (e.g., legal challenges).

Ms. Stovall emphasized the importance of full disclosure with annotation to allow interpretation and informed decision-making. Her organization, the National Coalition of Cancer Survivors, is working with providers in the spirit of offering "carrots" rather than "sticks." They also are working with large purchasers on patient-friendly displays of information. Ms. Stovall discussed the implications for consumers of not having good evidence regarding medical care. In the absence of evidence on the benefits of high-dose chemotherapy and bone marrow transplantation, providers offered and consumers demanded access to the procedure. Subsequent clinical trial data suggest that the procedure offers no substantial benefit over conventional therapies in most situations. Consumers often make choices based on the trust they have in their doctor and, increasingly, based on advice from consumer groups.

Purchasers' Perspectives

Mr. Bradley described employers' urgency to ensure that employees obtain optimal care and the ethical, legal, and fiduciary responsibility that companies have to ensure that their beneficiaries receive high-quality care. A group of employers has organized the Leapfrog Group to

apply the best available measures in purchasing decisions. At present, volume measures are felt to be ready and may serve to "stir the pot" to facilitate quality improvement. When provider specific volume data are publicly disclosed, some concerns have arisen regarding liability, especially if the inference is made that low volume reflects poorer outcomes.

Dr. Clauser suggested that his agency, the Health Care Financing Administration (HCFA), could use volume data to target interventions of its peer review organizations (PROs), which operate at the state level, to ensure the quality of care for Medicare beneficiaries. More useful, however, would be information on processes of care that underlie the relationship to affect outcomes. Also, establishing the relationship of volume of services to outcomes of particular relevance to the Medicare population, such as functional status, would be valuable. In Dr. Clauser's view, research is also needed to establish whether the volume–outcome relationship extends to nonsurgical interventions and to aspects of chronic care. HCFA is funding research to better understand the underpinnings of the volume–outcome relationship. A 13-state prospective outcome assessment study examining carotid endarterectomy and a number of PRO quality improvement initiatives in surgery safety and outcomes may guide Medicare decisions.

Dr. Clauser cautioned that volume–based selective referral may run counter to patient preferences and could narrow the choice of providers, a quality highly valued by consumers. In a dynamic market, it may also be difficult to apply volume thresholds. As part of HCFA's transplantation certification program, volume is one criterion, but volume can change quickly when, for example, a hospital's transplant team moves to another facility. Dr. Clauser pointed out that a monitoring program to keep track of changes in the marketplace requires an infrastructure that is expensive to maintain.

Dr. Clauser suggested that data about volume alone may be insufficient to shift care to higher-volume providers. The HCFA experience with centers of excellence for coronary artery bypass surgery suggests that it is very difficult to change physician referral patterns. When centers were able to advertise on the basis of their designation as a center of excellence, there was very little impact on referrals. Dr. Clauser concluded with a request that the research community provide answers to what volume means so that consumers can interpret the information that is increasingly being made available to them.

Managed Care Perspective

Dr. Isham pointed out the potential value of examining the relationship between volume, outcome, and price of procedures performed outside the usual reimbursement system, using a case study approach. There could, for example, be lessons relevant to cataract surgery in the recent surge in outpatient elective laser eye surgery. This procedure has expanded rapidly through high-volume centers' offering the procedure at relatively low cost. Dr. Isham suggested that physicians and hospitals could be required to maintain a logbook of their experience, similar to the logs that pilots are required to keep. Such a log could be made publicly available and could also serve as a tool for self-evaluation or as the basis of certification or accreditation. Dr. Isham raised concerns regarding the emergence of provider cartels in some geographic areas and also about practical applications at the market level (e.g., implications for rural areas).

Dr. Ho recounted experience at Pacificare Health Systems suggesting that disclosure to consumers of information on quality can affect consumer choice, improve performance, and reduce practice variation. Pacificare will consider integrating information about volume into its members' report card program. Twice a year, medical groups are ranked and findings are posted on the Internet (www.pacificare.com) and in the provider directory sent to each member. The

reports show the top-ranking groups (i.e., the top 10 percent) according to their performance in 32 categories related to clinical care, service, and administrative functions. The breadth of the rankings responds to the observation that consumers use different information in choosing providers. Public reporting has been associated with significant net increases in membership within the highly ranked groups and a reduction in variation in provider performance.

In considering the volume–outcome relationship as it applies to quality improvement, Dr. Ho expressed some concerns about unintended consequences of its application:

- The focus on high-volume providers may prove a distracting priority. Improvements might be more readily achieved through reducing fragmentation of services, better organization of systems of care, focus on prevention, the institution of standardized performance metrics, and an "upstream," population-based focus on health and disease management.
- Hospitals with a high volume of a certain procedure may use that as "brand leverage" to misrepresent their general experience.
- High-volume hospitals may achieve too much contractual leverage, giving rise to "cartels" and price inflation.
- More may not always be better: could there be a quality decrement with very high volume? Might inappropriate utilization be stimulated by volume-based rewards?
- Procedures counting toward high volume may not always be indicated or appropriate.
- More than a consensus on volume is needed. There could be a significant provider and/or consumer backlash if selective referral is widely implemented, unless all industry stakeholders demonstrate agreement.

Provider Group Perspective

Dr. Nielsen described the overarching goals of the health care system—improving health status, access to care, and the coordination and continuity of care—and suggested that information that furthers these goals needs to be disseminated to all stakeholders in a form that is accurate, reliable, and easily understood. In his view, volume as a measure of quality is a "blunt-edged tool." Given the limited state of research in this area, he pointed out that mortality may be improved with volume, but this is not necessarily true for other important outcomes of care (e.g., functional status, quality of life). In addition, higher volumes do not necessarily mean that there are decreases in associated morbidities such as infections or that certain critical steps in the process of care are utilized. There is the potential for misuse of such information, for example, misrepresentation in advertising. Furthermore, perverse incentives to do more procedures may well be created with the establishment of volume thresholds that could contribute to inappropriate care being rendered. There is also a potential for consumers to misinterpret the information and make decisions about care that are unwarranted (e.g., traveling great distances to tertiary care centers for simple procedures). Dr. Nielsen concluded that volume should not be used alone but should instead be supplemented with ancillary information and disclosed with all of its caveats.

Research Perspective

Dr. Fraser from the Agency for Healthcare Research and Quality noted that much of the research conducted to date has examined the clinical processes of care that underlie the volume–outcome relationship, but it has not yet addressed how systems of care might mediate this relationship. A study of the relationship between volume and outcome in acute myocardial infarction, for

example, showed that the underuse of beta-blockers in low-volume facilities helped explain the significant volume–outcome finding. Knowing what systems of care are associated with the appropriate use of beta-blockers in high-volume hospitals is critical to implementing this finding. How the volume–outcome relationship varies by community or health care market, and how the relationship might vary within hospital and by procedure, has pragmatic implications.

AHRQ's Healthcare Cost and Utilization Project (HCUP) is a potential resource for conducting studies of both hospital and physician volume. The HCUP State Inpatient Database includes more than half of all U.S. community hospital discharges (discharges from 22 states). A set of risk-adjusted quality indicators has been developed that could be used as benchmarks for health services researchers. Procedure-specific data sets could be produced for researchers with a specialty interest. Market-specific data, even if qualitative, would be of interest to purchasers to assist in their decision-making. Exceptions to the expected relationship—for example, high-volume, low-quality hospitals, could be explained by a low pricing strategy adopted by a particular hospital. Dr. Fraser suggests that research is also needed on the process and outcome of the purchasing strategy itself—its consequences on the quality of care and its market impact. For implementation, we need to know what kinds of data are most useful to purchasers and how employees can best use this information.

Dr. Fraser also pointed out that consumer use of information about volume of services is limited by the fact that it is usually employers, not employees, who choose a health plan. Most employers offer only one or two plans to their employees, and a minority offers plans that provide a great deal of choice at the point of care.

Dr. Berwick led off the discussion by asking whether the potential for unintended consequences of public disclosure of volume–outcome data was sufficient to delay or limit public disclosure.

Dr. Simone pointed out that consumers are bombarded with medical information, much of it through advertising. He suggested that consumers need explanation but that it cannot be too complicated.

Dr. Chassin suggested that there are different thresholds for different actions. In his view, data on volume should be publicly available with appropriate explanation. All states have some kind of hospital discharge database, and probably two-thirds of states have publicly available hospital data. Purchasers could take this information and use it for decision-making and/or could make the information public. New York, through its CON process, has utilized evidence on volume to achieve regionalized cardiac surgery. New York has 32 hospitals offering cardiac surgery, while California has 125 such programs. The net effect is that close to 80 percent of patients having bypass surgery in New York obtain it in hospitals doing more than 500 procedures a year. In contrast, less than one-half of patients in California undergo surgery in such high-volume settings. The New York regulatory process exists alongside a voluntary program, the cardiac surgery reporting system. With the adoption of these programs, there has been a significant decline in the mortality associated with cardiac surgery in New York, one that exceeds the general decline across the country. There is no evidence that the procedures are overused. In fact, estimates of rates of overuse of angioplasty, angiography, and bypass surgery are very low—4 percent for percutaneous cardiac angioplasy and 2 percent for bypass surgery. Furthermore, estimates of rates of underuse are similar to those observed in California. Lastly, access to these procedures in New York and California is similar—patients need not travel farther for high-volume care. The combined regulatory and state-sponsored voluntary program works, but difficult political battles had to be won before it could be implemented

Ms. DelBanco, from the Leapfrog Group, felt optimistic that its evidence-based outcome referral program would motivate change. Dr. Hillner added that data systems tend to get better with use (e.g., Health Plan Employer Data and Information Set).

Dr. Hannan again stated the limitations of making inferences about the performance of individual providers with administrative data (e.g., statewide hospital discharge data). Although these data can be used to say that, in the aggregate, outcomes for certain procedures are worse in low-volume hospitals, the performance of individual small-volume providers cannot be predicted with confidence because of the instability of small numbers (e.g., a few deaths can greatly skew annual rates). Similar problems arise when evaluating rare conditions or procedures (e.g., surgery for pancreatic cancer). Dr. Hannan recommends against using volume data alone to judge the performance of individual providers. Instead, volume data, if used, should be coupled with risk-adjusted outcome data from administrative databases or used in concert with crude mortality rates. The use of a combination of sources is preferable to using any nonclinical source alone. This will minimize referrals to the numerous high-volume hospitals with poor outcomes. He also pointed out that referrals are generally made to physicians practicing within hospitals, and ideally, one would have quality data at both levels of care.

Dr. Hannan suggests that the most important use of volume data is to identify processes and structures of care that distinguish high- and low-volume providers and that predict outcomes. Findings can then be used to enhance the performance of providers. In an examination of the surgeon volume–outcome relationship among patients undergoing carotid endarterectomy in New York State, volume effects disappeared when type of surgeon was controlled for in the analysis (i.e., vascular surgeons had lower risk-adjusted mortality rates than either general surgeons or neurosurgeons). Furthermore, the use of more appropriate medications and surgical techniques by vascular surgeons explained their mortality advantage. This study illustrates that volume, which may appear to be an important driver of quality, may not be significant once processes and structures of care are taken into account. The ability to tease out the relative contributions of volume, process, and structure of care were possible in this study only because comprehensive data are systematically collected in New York as part of a carotid endarterectomy registry.

For procedures performed infrequently, Dr. Dudley pointed out that it is very difficult to measure quality, so proxies such as volume may have to do. For some conditions with low overall volume (e.g., esophageal cancer), he suggests that it may be impossible to refer selectively based on direct measurement of quality (i.e., physician or hospital-specific outcome data). Here, one could base selective referral on minimum volume standards. For conditions in which high volume has been shown to improve outcome and for which case loads are large enough to support outcomes measurement (e.g., CABG), it is also feasible to base selective referral on outcome, rather than volume.

Dr. Hertzer, a vascular surgeon at the Cleveland Clinic, suggested that prospective outcome assessment appears to be successful in improving physician performance. In an HCFA-sponsored demonstration of carotid endarterectomy, at least some participating areas found improvements in processes of care and outcomes associated with such an assessment.

Dr. Simone discussed the pediatric oncology model as a potential way to reorganize care to achieve optimal outcomes. Most children with cancer are enrolled in clinical trials, irrespective of the site of care. They are cared for in community settings, but according to research protocols that guarantee the use of best practices. All pediatric oncologists contribute outcome data for assessment. Mentoring relationships exist between high- and low-volume hospitals.

WHERE DO WE GO FROM HERE?

Dr. Simone concluded the workshop by asking each participant to consider the day's proceedings and to assess what policy and research actions were warranted.

Many workshop participants noted that volume may be the best available proxy indicator of quality for certain conditions, but efforts should be made to accompany volume information with other quality indicators and with explanatory information. Furthermore, several participants suggested that when research confirms a volume–outcome link, this information should be disclosed to the public to support health care decision-making. In making such disclosures, however, the limitations of the data and how the information should be interpreted must be clear for its intended audience. Public release of data on volume may motivate providers to report process and outcome data, thereby laying the ground for the next generation of health care quality measures. When better measures of quality of care than volume are developed, they should replace the volume measures

Many people could eventually be affected by policies that concentrate care in higher-volume settings (or that avoid low-volume settings) because many procedures for which a volume–outcome relationship has been established are conducted by low-volume providers. Many barriers to implementing volume-based policies were recognized, and implementation was viewed as especially challenging in rural areas. Residents of rural areas, for example, may not want to travel great distances for care and may not be able to afford to do so. Consequently, some workshop participants suggested that initial implementation efforts be focused in urban areas with dense concentrations of hospitals, where success is more likely.

Purchasers were recognized as having taken the lead in applying the evidence regarding volume. The Leapfrog Group, for example, is exploring an evidence-based referral program within the health plans with which it contracts. For conditions or procedures for which the research evidence strongly points to a volume–outcome relationship, a next step for action might be achieving consensus, perhaps with the help of representatives of specialty providers, on acceptable volume thresholds for selected procedures. To apply the volume–outcome findings more broadly, further research is needed on many conditions and on many health care outcomes for which data are now insufficient.

Workshop participants suggested a wide-ranging set of research topics —from policy research and demonstration programs to basic methodological research—to better understand why the relationship between volume and outcome exists and how best to implement policies to improve care.

Policy Research and Demonstration Programs

- Make the "business case" for implementing a volume standard when assessing quality of care, and examine the implications of policies that would effectively stop certain procedures from being performed in low-volume hospitals.
- Evaluate current efforts to implement the volume–outcome finding (e.g., the Leapfrog Group's evidence-based referral program).
- Plan demonstrations with quasi-experimental designs to evaluate the impact of applying different quality-of-care measures to quality improvement programs (e.g., volume versus risk-adjusted administrative data plus mortality data).

Consumer-Focused Research

- Examine consumer interest in, and interpretation of, volume as an indicator of health care quality relative to other measures.
- Evaluate the relationship between exposure to volume quality indicators and consumer health care decision-making.
- Examine dissemination strategies to maximize consumer access to and use of quality information (e.g., determine which sources of information are trusted).

Implementation Related Research

- Determine thresholds with which to define high volume for each condition or procedure (e.g., using focus groups of members from specialty societies, syntheses of extant literature).

New Areas of Research

- Examine the volume–outcome relationship for chronic conditions and nonsurgical procedures.
- Examine factors that mediate volume–outcome relationships when found.

Methodological Development

- Examine outcomes other than mortality (e.g., functional status, quality of life), including longer-term outcomes.
- Include potential intervening variables (e.g., processes and systems of care) in volume–outcome studies.
- Integrate social science methods (e.g., sociology, medical anthropology, systems analysis) into volume–outcome research.
- In studies with mortality as an end point, examine the time of death to help explain the cause of death.
- Develop procedure-condition-specific risk-adjustment tools.

Health Services Research Data Infrastructure

- Develop condition-procedure-specific, prospective, population-based clinical databases and registries (e.g., New York's cardiovascular surgery database).
- Develop chronic disease databases that cover both hospital and outpatient care.
- When clinical databases are available to serve as a "gold standard," evaluate the sensitivity and specificity of volume as a quality indicator.
- Selectively add key clinical risk factors and process data into administrative databases.
- Include more reliable identifiers of physicians in federal and state administrative databases.

Workshop participants cited several potential sources of research support including AHRQ, health care purchasers, health plans, and provider groups, and suggested that public–private research partnerships might greatly facilitate meaningful research efforts.

APPENDIX A
Workshop Agenda

INSTITUTE OF MEDICINE

Division of Health Care Services,
Committee on Quality in Health Care in America

National Cancer Policy Board

May 11, 2000

OBJECTIVES

1. Review evidence of the relationship between volume of services and health-related outcomes.
2. Discuss methodological issues related to the interpretation of the association between volume and outcome.
3. Assess the applicability of volume as an indicator of quality of care.
4. Identify research needed to better understand the volume-outcome relationship and its application to quality improvement.

AGENDA

9:00 a.m.–9:30 a.m.	Welcome and introduction: Diana Petitti Opening remarks: John Eisenberg

Session 1

9:30 a.m.–12:15 p.m.	Overview of the Volume-Outcome Relationship Opening remarks: Mark Chassin
9:30 a.m.–10:30 a.m.	Presentation of background paper: "Synthesis of Evidence from the Literature," Ethan Halm
10:30 a.m.–10:45 a.m.	Coffee break
10:45 a.m.–12:15 p.m.	Roundtable discussion Moderators: Mark Chassin and Jane Sisk

- Is the association consistent across conditions/procedures/services?
- What factors might explain the association?

- What measurement issues need to be addressed (e.g., defining volume, thresholds/cutpoints, units of analysis)?
- Does the evidence support applying volume as a quality-of-care measure?
- What additional research is needed?

Reactors/discussants: Edward Hannan, Norman Hertzer, Colin Begg, Bruce Hillner, Arnold Milstein

12:15 p.m.–1:15 p.m. Lunch

Session 2

1:15 p.m.–3:30 p.m. Policy implications
Opening remarks: Diana Pettiti

1:15 p.m.–1:45 p.m. Presentation of background paper—"When and How Could Volume Be Used as an Indicator of Quality Care?" Adams Dudley

1:45 p.m.–3:30 p.m. Roundtable discussion
Moderators: Don Berwick and Robert Galvin

- What level of evidence is needed to support policies?
- What are the strengths and weaknesses of alternative implementation strategies?
- What are the potential barriers to applying volume as an indicator of quality (e.g., access in rural areas, patient preferences for care close to home)
- What policy-relevant health services research is needed?

Reactors/discussants:

- Consumer perspectives: Art Levin and Ellen Stovall
- Purchaser perspectives: Bruce Bradley, Steven Clauser, and Irene Fraser
- Managed care: George Isham and Sam Ho
- Provider groups: Don Nielsen

Session 3

3:30 p.m.–4:00 p.m. Where do we go from here?
Moderator: Joseph Simone

Wrap-up: Discussion of findings and research needs

4:00 p.m. Workshop Adjourns

PARTICIPANTS

Colin Begg
Memorial Sloan-Kettering Cancer Center
New York, NY

Donald M. Berwick
Institute for Healthcare Improvement
Boston, MA

Richard Bae
University of California
San Francisco, CA

Bruce Bradley
General Motors
Detroit, MI

Mark R. Chassin
The Mount Sinai School of Medicine
New York, NY

Steven Clauser
Health Care Financing Administration
Baltimore, MD

Jan De la Mare
Agency for Healthcare Research and Quality
Rockville, MD

Suzanne DelBanco
The Leapfrog Group
Washington, DC

R. Adams Dudley
UC San Francisco
San Francisco, CA

John Eisenberg
Agency for Healthcare Research and Quality
Rockville, MD

Irene Fraser
Agency for Healthcare Research and Quality
Rockville, MD

Robert Galvin
General Electric Company
Fairfield, CT

Ethan Halm
The Mount Sinai School of Medicine
New York, NY

Edward Hannan
State University of New York at Albany
Rensselaer NY

Norman Hertzer
Cleveland Clinic
Cleveland, OH

Bruce Hillner
Virginia Commonwealth University
Richmond, VA

Sam Ho
Pacificare Health Systems
Santa Anna, CA

George J. Isham
HealthPartners
Bloomington, MN

Clara Lee
The Mount Sinai School of Medicine
New York, NY

Arthur Levin
Center for Medical Consumers
New York, NY

Arnold Milstein
William M. Mercer, Inc
San Francisco, CA

Peggy McNamara
Agency for Healthcare Research and Quality
Rockville, MD

Don Nielsen
American Hospital Association
Chicago, IL

Diana Petitti
Kaiser Permanente of Southern California
Pasadena, CA

Joseph Simone
Huntsman Cancer Foundation and Institute
Salt Lake City, UT

Jane Sisk
Mount Sinai School of Medicine
New York, NY

Ellen Stovall
National Coalition for Cancer Survivorship
Silver Spring, MD

IOM and NCPB Staff

Janet M. Corrigan, Program Director, IOM
Maria Hewitt, Senior Program Officer, NCPB
Ellen Johnson, Administrative Assistant, NCPB
Kelly C. Pike, Senior Project Assistant, IOM

APPENDIX B
Rosters

COMMITTEE ON QUALITY OF HEALTH CARE IN AMERICA

William C. Richardson, Ph.D. (*Chair*)
President and CEO
W.K. Kellogg Foundation
Battle Creek, MI

Donald M. Berwick, M.D., M.P.P.
President and CEO
Institute for Healthcare Improvement
Boston, MA

J. Cris Bisgard, M.D., M.P.H.
Director, Health Services
Delta Airlines, Inc.
Atlanta, GA

Lonnie R. Bristow, M.D.
Past President
American Medical Association
Walnut Creek, CA

Charles R. Buck, Sc.D.
Program Leader, Health Care Quality and Strategy Initiatives
General Electric Company
Fairfield, CT

Christine K. Cassel, M.D.
Professor and Chairman
Department of Geriatrics and Adult Development
Mount Sinai School of Medicine
New York, NY

Mark R. Chassin, M.D., M.P.H., M.P.P.
Professor and Chairman, Department of Health Policy
The Mount Sinai School of Medicine
New York, NY

Molly Joel Coye, M.D., M.P.H.
Pricewaterhouse Coopers Senior Fellow
Institute for the Future
Menlo Park, CA

Don E. Detmer, M.D.
Professor of Medical Education in Health Evaluation Sciences
University of Virginia
Charlottesville, VA

Jerome H. Grossman, M.D.
Chairman and CEO
Liongate Management, LLC
Boston, MA

Brent James, M.D.
Executive Director
Intermountain Health Care
Institute for Health Care Delivery Research
Salt Lake City, UT

David McK. Lawrence, M.D., M.P.H.
Chairman and Chief Executive Officer
Kaiser Foundation Health Plan, Inc.
Oakland, CA

Lucian Leape, M.D.
Adjunct Professor of Health Policy
Department of Health Policy Management
Harvard School of Public Health
Boston, MA

Arthur Levin, M.P.H.
Director, Center for Medical Consumers
New York, NY

Rhonda Robinson-Beale, M.D.
Executive Medical Director of Managed Care Management and Clinical Programs
Blue Cross Blue Shield of Michigan
Southfield, MI

Joseph E. Scherger, M.D., M.P.H.
Associate Dean for Clinical Affairs
University of California, Irvine
College of Medicine
Irvine, CA

Arthur Southam, M.D.
Partner
2C Solutions
Northridge, CA

Mary Wakefield, R.N., Ph.D.
Director, Center for Health Policy and Ethics
George Mason University
Fairfax, VA

Gail L. Warden
President and Chief Executive Officer
Henry Ford Health System
Detroit, MI

IOM Staff

Janet Corrigan, Project Director
Molla Donaldson, Project Codirector
Linda Kohn, Project Codirector
Kelly Pike, Senior Project Assistant

NATIONAL CANCER POLICY BOARD
INSTITUTE OF MEDICINE/NATIONAL RESEARCH COUNCIL

Arnold J. Levine, M.D. (*Chair*)
President
The Rockefeller University
New York, New York

Joseph Simone, M.D. (*Vice Chair*)
Medical Director
Huntsman Cancer Foundation and Institute
Salt Lake City, UT

Ellen Stovall (*Vice Chair*)
Executive Director
National Coalition for Cancer Survivorship
Silver Spring, MD

Diana Petitti, M.D. (*Vice Chair*)
Director, Research and Evaluation
Kaiser Permanente of Southern California
Pasadena, CA

Tim Byers, M.D., M.PH.
Professor of Epidemiology
Program Leader, Clinical Cancer Prevention Control
University of Colorado Health Sciences Center
Denver, CO

Vivien W. Chen, Ph.D.
Epidemiology Section Chief and Professor
Louisiana State University Medical Center
New Orleans, LA

Susan Curry, Ph.D.
Director
Center for Health Studies
Group Health Cooperative of Puget Sound
Seattle, WA

Norman Daniels, Ph.D.
Professor of Philosophy
Tufts University
Newton, MA

Kathleen Foley, M.D.
Chief, Pain Service
Department of Neurology
Memorial Sloan-Kettering Cancer Center
New York, NY

Tom Kelly, M.D.
Boury Professor and Chairman
Department of Molecular Biology and Genetics
Johns Hopkins University School of Medicine
Baltimore, MD

Mark McClellan, M.D., Ph.D.
Assistant Professor of Economics
Stanford University
Stanford, CA

William McGuire, M.D.
CEO, UnitedHealth Group
Minnetonka, Minnesota

John Mendelsohn, M.D.
President
M.D. Anderson Cancer Center
University of Texas
Houston, TX

Monica Morrow, M.D.
Professor of Surgery
Comprehensive Breast Program
Northwestern University
Chicago, IL

Nancy Mueller, Sc.D.
Harvard School of Public Health
Department of Epidemiology
Boston, MA

Pilar Ossorio, Ph.D., J.D.
Assistant Professor of Law and Medical Ethics
Associate Director for Programming, Center for the Study of Race and Ethnicity in Medicine
University of Wisconsin Law School
Madison, WI

Cecil B. Pickett, Ph.D.
Executive Vice President
Discovery Research
Schering-Plough Research Institute
Kenilworth, NJ

John Seffrin, Ph.D.
Chief Executive Officer
American Cancer Society
Atlanta, GA

Sandra Underwood
ACS Oncology Nursing Professor
University of Wisconsin
School of Nursing
Milwaukee, WI

Frances Visco
President
National Breast Cancer Coalition
Washington, D.C.

Susan Weiner, Ph.D.
President
The Children's Cause
Silver Spring, MD

APPENDIX C

How Is Volume Related to Quality in Health Care? A Systematic Review of the Research Literature

Ethan A. Halm, M.D., Ph.D.; Clara Lee, M.D., M.P.P.; and
Mark R. Chassin, M.D., M.P.P., M.P.H.
Department of Health Policy, Mount Sinai School of Medicine

EXECUTIVE SUMMARY

Measuring and understanding the relationship between volume and outcome in the delivery of health services has engaged the attention of an increasing number of researchers since the 1980s. More recently, health care purchasers have begun to debate whether and how they should make use of these research findings in their relationships with health plans and individuals.

We conducted a systematic review of the published research literature to assess the methodological quality of the research and to understand what is known about how volume and outcome are associated, by what causal pathways they might be linked, and how generalizable the research findings are. We developed an 18-point scale to rate the methodological rigor of each study. The scale was based on a conceptual model that identified a series of critical factors that need to be addressed to fully delineate the relationship between volume and quality of care. These factors include: adequacy of risk-adjustment for differences in severity and comorbidity, sample size, unit of analysis, assessment of specific clinical processes of care, organizational processes, skills and experience of the various relevant clinicians, appropriateness of patient selection, and the spectrum of important clinical outcomes.

We identified published, English-language studies by electronic searches of MEDLINE, by consulting with experts, and by hand searching the reference lists of identified studies. We excluded studies with patient cohorts treated prior to 1980, samples that were not community or population-based, those in which a health outcome was not the dependent variable, and those in which volume was not an independent variable. This monograph reports the results of our review of 8 conditions and procedures: coronary artery bypass graft (CABG) surgery, pediatric cardiac surgery, carotid endarterectomy, abdominal aortic aneurysm repair, cancer surgery, percutaneous transluminal coronary angioplasty (PTCA), acute myocardial infarction, and acquired immuodeficiency syndrome (AIDS).

Overall, 88 studies were included in this review. The median quality score was 8, less than 50% of the maximum possible score of 18. The highest score assigned (to 1 study) was 13, 72% of the maximum possible score. All of the remaining studies scored 11 or less. The principal omissions in the published research, which led to lower quality scores, included:

1. Only 16 considered the independent effects of both physician and hospital volume;

2. Some (n = 24) used clinical data to adjust for differences among patients in severity and comorbidity, but only 4 reported statistically robust models;
3. Few (n = 4) adjusted for differences in specific processes of care; and
4. Few studies (n = 2) addressed appropriateness of patient selection.

Several conclusions emerged from this review. There can be little doubt that for a wide variety of surgical procedures and medical conditions higher volume (whether assessed by hospital or by physician) is associated with better health outcomes. Although a publication bias against negative findings appearing in print cannot be excluded, the uniformity with which the published research documents or confirms the existence of the association is compelling. Fully 77% of the studies we reviewed found statistically significant associations between higher volume and better outcomes. The remaining 23% did not find statistically significant relationships. No study demonstrated a statistically significant association in the opposite direction. Finally, all 16 of the studies with the highest quality scores found statistically significant associations.

Overall, however, the methodological rigor of the studies was modest, limiting generalizability and leaving many questions about the nature and causes of the association unanswered. Studies of the same procedure or condition typically employed widely varying definitions of high and low volume, precluding definitive conclusions about the nature of its relationship to outcome. We do not know, for example, whether a volume threshold exists—or more than one—above which outcomes are better but do not continue to improve with further volume increases. Is there a threshold effect for physician or hospital volume? Both? Neither? Very few studies addressed the question of whether the only relevant volume measure is that of the specific procedure or condition under study. Thus, we do not know the extent to which experience with procedures (or conditions) that are closely related affects outcome.

Available research does not shed much light on what specific factors explain outcome differences between high and low volume providers. No longitudinal studies (and very few of any design) address the important question of how much of the variability, especially among low-volume providers, is due to chance. Few investigations have assessed differences in specific clinical processes of care, especially those known to affect outcomes. One intriguing exception showed that about one third of the mortality difference between high- and low-volume hospitals for acute myocardial infarction could be attributed to more frequent use of proven-effective medications at high-volume hospitals. We believe that this direction is the most promising avenue for future research: a detailed understanding of those critical clinical and organizational processes that affect outcomes and that differentiate high- and low-volume hospitals and physicians. Such knowledge could provide a blueprint for improvement.

These research findings have several potential implications for public policy in health care. Public policy might be directed at informing consumers and referring physicians about the nature of the relationship between volume and outcome or at improving outcomes, or both. Data on volume of procedures by hospital are publicly accessible in many states through hospital discharge abstract databases. Many of these databases also permit identification of treating physicians. Making these data widely available is one potential way of informing consumers and clinicians. Some of the challenges such efforts would face are the difficulties of explaining what the data mean (and what they do not), the technical challenges of carefully identifying conditions and procedures for which research is conclusive that higher volume is associated with better outcomes (and at what threshold), and the need to update the data regularly.

Regulation in some states has limited the number of low-volume hospitals for some specific services, but such programs have become increasingly politically unpopular and require significant infrastructure to support. Recent actions by private employers suggest that some purchasers may be prepared to take an active role in steering patients toward high-volume hospitals, using incentives directed either at managed care plans or at employees themselves. No data exist on which to base predictions of the impact of such efforts. Available data suggest that significant numbers of patients do not use data on outcomes or volume where they exist to choose where to receive care.

We believe that the most promising use of the research findings linking volume and outcome is to fuel quality improvement. To the extent ongoing research is successful in demonstrating what specific clinical and organizational processes of care are associated with improved outcomes, this knowledge can be put to direct and immediate use to improve hospital and physician care. Public policy can accelerate this improvement by supporting the necessary research, by fostering and evaluating demonstration programs to implement it, and by disseminating successful models.

HOW IS VOLUME RELATED TO QUALITY IN HEALTH CARE? A SYSTEMATIC REVIEW OF THE RESEARCH LITERATURE

> "We must formulate some method of hospital report showing as nearly as possible what are the results of the treatment obtained at different institutions. This report must be made out and published by each hospital in a uniform manner, so that comparison will be possible. . . . A set of statistics had been prepared comparing the mortality at a certain semi-private hospital of 200 beds with that of four of the best general hospitals in America, having a total of 1,200 beds. These statistics were obtained from published reports. They clearly showed that the semi-private hospital not only did more operations, but that the mortality was much lower, especially in some of the more difficult branches of surgery."

Introduction and Historical Context

This author goes on to speculate about what factors might have produced such an apparently anomalous result—that a small hospital might have better outcomes than larger, more prestigious institutions. Among the factors he cited in addition to the volume of operations were the skill of the surgeons, the comorbid conditions of the patients, the financial condition of the hospitals, and the way in which surgical departments were organized to triage particular patients to particular surgeons. Data to answer these questions continue to prove elusive, some 86 years after a Boston surgeon, Ernest Avery Codman, first posed them. (Codman 1914)

Codman's questions lay dormant for more than 50 years, until the National Halothane Study, undertaken in the 1960s to assess possible hepatotoxic effects of halothane, unexpectedly reported a 24-fold variation in operative mortality among participating U.S. hospitals. (Moses 1968) This finding led to a series of further investigations. (Stanford 1976, Flood 1979, Flood 1984a, Flood 1984b) More recently, beginning with the work of Luft and colleagues (Luft 1979, Bunker 1982, Luft 1987), a large body of research has focused on measuring and explaining the relationship between patient outcomes and the volume of specific health services provided by hospitals and physicians.

Many studies have documented that higher volume is associated with better outcomes for a variety of different surgical procedures and medical conditions. Two principal hypotheses have

been advanced to explain these relationships: 1) that physicians (and hospitals) develop more effective skills if they treat more patients ("practice makes perfect") or 2) physicians (and hospitals) achieving better outcomes receive more referrals and thus accrue larger volumes ("selective referral"). (Flood 1984a, Flood 1984b, Luft 1987)

Although a few studies have found some indirect evidence that is consistent with either or both of these hypotheses, we believe a reconceptualization is warranted of the model that describes the potential mechanisms by which volume and outcome might be related in health care. This reconceptualization is prompted both by several inconsistencies in the older hypotheses and by evidence from recent research. Although the "practice makes perfect" hypothesis has some intuitive appeal, extant research has been unable to demonstrate what skills or practices improve as hospitals or physicians gain experience with particular kinds of patients and why they are uniquely related to volume. For the "selective referral hypothesis" to be an important explanatory factor, the outcomes of different physicians, surgeons, and hospitals would have to be widely known for referring physicians (if not patients) to use in making treatment recommendations. But we know that such data are not readily available and cannot be the basis for very many patient referrals. Even where they do exist, physicians report that they do not use them to make referrals. (Schneider 1996, Hannan 1997) Nor is there evidence that patients either shun hospitals with well-publicized poor outcomes or flock to those with good ones. (Chassin 1996, Schneider 1998) Further, both of these hypotheses focus on explaining how particular hospitals or physicians might have achieved high volume and good outcomes over time. Neither offers much help in explaining the causes of differences in outcomes among hospitals or physicians with different volumes or in suggesting how the poorer performers might improve. Some recent evidence documenting systematic differences among high- and low-volume providers in the use of specific interventions points to a different set of explanatory factors. (Thiemann 1999)

We, therefore, begin this systematic review of the research evidence linking volume and outcome in health care by describing a new conceptual model. In subsequent sections of this monograph, we describe how we used this model to develop criteria to judge the quality of the published research we reviewed. We then define our literature search strategies, inclusion and exclusion criteria, and methods for reviewing and rating included studies. We discuss the results of the review separately for each included topic area and then discuss generalizable findings and issues. Finally, we discuss the potential policy implications of the findings.

Conceptual Framework

Figure 1 displays the factors we believe are likely to be the most potent explanatory variables in understanding how volume of services is related to health outcomes. Patients are selected to receive surgical or medical services in a variety of ways. Some have little time or choice (e.g., those suffering acute myocardial infarction or ruptured aortic aneurysms). The opposite is true for many surgical procedures, especially for the vast majority of elective operations. We know little about the relationship between volume and patient selection. Are patients of high-volume hospitals or physicians likely to be more or less appropriate candidates for surgery than patients of low-volume providers? A single study documented that high-volume surgeons were more likely to perform inappropriate carotid endarterectomies than low-volume surgeons were. (Brook 1990b) Few investigations have addressed this important dimension of quality.

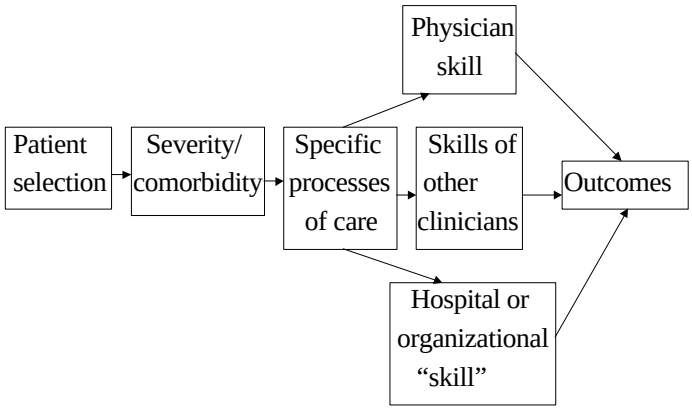


Figure 1 Conceptual Framework: How Could Volume Affect Quality?

More attention has been focused on identifying the key variables that define the severity of patients' presenting illness and the comorbid conditions that also affect the outcomes of treatment. How well these factors are measured and taken into account (i.e., how well the data on outcomes are "risk-adjusted") are crucial components of the validity of outcome comparisons between high- and low-volume providers. If high-volume surgeons or hospitals treat patients who are systematically sicker (or healthier) than their low-volume counterparts, comparisons of their outcomes are not valid without scientifically rigorous risk adjustment.

Once the patient population undergoing treatment is adequately characterized, it is then necessary to consider exactly what treatments the patients receive. Volume of services cannot directly produce better outcomes. If volume is related to outcome, that association must be expressed by differences in the components of care or in the skill with which treatments are provided. For some conditions or procedures, specific component treatments (processes of care) have been demonstrated to improve outcomes. For example, it is essential to know whether patients undergoing treatment for acute myocardial infarction received therapies designed to achieve reperfusion (acute angioplasty or thrombolytic drugs in appropriate settings). Other effective processes for this condition include aspirin, beta blockers, and angiotensin converting enzyme (ACE) inhibitors, each with its own specific indications. Although data on effectiveness and efficacy for most surgical processes of care are lacking, it is equally essential that researchers investigate the extent to which carotid endarterectomies or coronary bypass operations are performed the same way at high- and low-volume centers and by different surgeons. Do high-volume surgeons employ different surgical techniques compared with low-volume surgeons? If they do different operations, we can characterize the differences and assess the extent to which these measurably different processes of care are associated with different outcomes.

If they do exactly the same operations but obtain different results, we may attribute such outcome differences to differences in skill (or "unmeasured process" differences). Several aspects of this issue are worth noting. First, if physician experience (volume) is a proxy measure for the attainment of a certain skill level that permits superior performance, it is important to consider how that experience is best measured. Does superior performance in carotid endarterectomy or aortic aneurysm surgery derive from experience with just those procedures, or is the volume of other major vascular surgical procedures important also? This issue may be termed the "volume of what" question.

Second, it is important also to identify exactly what skills are related to outcomes. In surgical procedures, specific physical skills may be vital—for example, the speed and dexterity with which a carotid artery is dissected, minimizing trauma to adjacent tissues. Performance of tasks like these may improve as volume increases and may require a certain level of volume to achieve or maintain proficiency. The acquisition and maintenance of certain cognitive skills may also be important to producing good outcomes and related to volume. Recognizing uncommon presentations of medical conditions or anatomic variations in surgery and taking appropriate and timely action are examples. To the extent the physical and cognitive skills most crucial to outcomes can be characterized and specified, they may move from the realm of "unmeasured processes" to assume their proper role in the set of crucial and measurable processes that are integral to producing the best outcomes.

Third, is the experience of other key participants important? Is the volume of experience in a hospital's emergency department or coronary care unit important to its outcomes for patients with acute myocardial infarction? What about the experience of key consultants? Is the amount of experience of the team managing the cardiopulmonary bypass pump important in contributing to the outcomes of coronary artery bypass surgery? Fourth, what about "organizational skill?" Do high-volume hospitals adopt specific organizational strategies that are particularly effective in enhancing outcomes? Do they employ sophisticated computerized reminder systems or standardized protocols that reduce errors of omission or commission? Do teaching hospitals, which are often high in volume, have an advantage because house staff and fellows provide 24-hour physician coverage? The more complex the treatment process, the more likely it is that physician or surgeon skill will be only one of many important components of the full complement of effective care. Defining the relative contributions of all important contributors is an essential research challenge.

Criteria for Rating the Quality of Published Studies

Using this conceptual model as a guide, we developed a scoring system to assess the quality of the research studies included in our systematic review. The full list of criteria is attached as Appendix A. Our aim was to create a quantitative method of assessing the research design of the studies we reviewed such that higher scores would reflect increasing likelihood of the study's ability to discern generalizable conclusions about the nature and magnitude of the relationship between volume and outcome. The first four criteria assess various aspects of the patient sample used in the research. We assigned one point if the sample was representative of the general population of all patients who might receive the treatments under study. Thus, studies of managed care plan enrollees or Medicare beneficiaries were not considered representative. We assigned two points if the study included patients of 50 or more physicians and 20 or more hospitals. If only one of these criteria was met, we assigned one point. No points were assigned if neither criterion was met. In some studies authors reported the number of hospitals in their sample but not the number of treating physicians. In these cases we estimated the number of physicians by assuming it would be at least equal to the number of hospitals. The vast majority of these studies included hundreds of hospitals from administrative databases, so we estimated the number of physicians as 50 for scoring this criterion. If the total sample size was 1000 patients or more, we assigned one point. Because statistical power to detect significant relationships in logistic regression models depends more on the total number of adverse events represented in the sample than on total sample size (and because the various conditions and procedures in this lit-

erature have widely varying adverse event rates), we assigned 2 points if the total number of adverse events was greater than 100, one point if it was 21–100, and no points if it was 20 or less.

We assigned no points if the study assessed the relationship between outcome and either hospital or physician volume. If both were assessed separately, we assigned one point. If the joint relationships of hospital and physician volume were assessed independently in a multivariate analysis, we assigned 2 points. And if a study examined both of these and the volume of another important component of the care process, we assigned 3 points. If the appropriateness of patient selection was not addressed, we assigned no points. If appropriateness was measured, we assigned one point. If it was measured and taken into account in the analysis of the volume/outcome relationship, we assigned 2 points.

If volume was analyzed in only 2 categories, we assigned no points. If more than 2 categories were assessed or if volume was treated as a continuous variable, we assigned one point to credit a more sophisticated assessment of a possible dose-response relationship. In considering the various ways in which outcomes might be risk-adjusted, we assigned no points if no risk-adjustment at all was done. If data from insurance claims, hospital discharge abstract databases, or other sources of administrative data were used, we assigned one point. If data from clinical sources (e.g., medical records or prospectively designed clinical registries) were used for risk-adjustment, we assigned 2 points. If clinical data were used in a logistic regression model that demonstrated good calibration by the Hosmer-Lemeshow test and good discrimination (by a C-statistic of 0.75 or greater), we assigned 3 points.

If specific clinical processes of care were not measured, we assigned no points. If a single process was measured and its impact on risk-adjusted outcomes assessed, we assigned one point. If two or more such processes were measured and evaluated, we assigned 2 points. Finally, if death was the only outcome evaluated, we assigned no points. If other adverse outcomes in addition to mortality were assessed, we assigned 2 points.

Quality scores were summed across all 10 criteria for each study. The maximum possible total score was 18.

Literature Review Methods

We performed two electronic subject-based searches of the literature on MEDLINE (1966–1999). A professional reference librarian assisted us in the development of our search strategy. First, we selected twelve conditions and procedures for which volume and outcomes had been studied (coronary artery bypass, carotid endarterectomy, peripheral vascular surgery, trauma, transplant, hip surgery, knee surgery, liver resection, lung resection, cancer surgery, angioplasty, and myocardial infarction).

We then developed a list of search terms based on subject headings from articles known to be highly relevant to our topic and from the official indexing terms of the MEDLINE database. We performed multiple searches with combinations of these terms and evaluated the results of those searches for sensitivity and specificity, with respect to our topic of volume and outcomes. The search algorithm that yielded the greatest number of highly relevant articles combined the 12 conditions with the terms volume, utilization, frequency, statistics, and outcomes. In order to broaden our search to include articles on regionalization of care, we added another search that combined the 12 conditions with the term regionalization. (See Appendix B.)

We also performed MEDLINE searches on authors known to have published widely on the study topic, and we searched the Cochrane Collaboration Database for systematic reviews. In addition to performing electronic database searches, we consulted experts in the field for further

references. Finally, we reviewed the references cited by each article that was ultimately included. We did not hand-search any journals. This review was limited to the English-language research literature.

This monograph includes the findings of our review of eight procedures and conditions: coronary artery bypass graft (CABG) surgery, pediatric cardiac surgery, carotid endarterectomy, abdominal aortic aneurysm repair, cancer surgery, coronary angioplasty, acute myocardial infarction, and AIDS. Our electronic search also identified articles about volume and outcome in trauma, newborn intensive care, organ transplantation, hip surgery, knee replacement, neurosurgery, prostatectomy, cataract surgery, endocrine surgery, and general surgical procedures for benign disease, but we did not include them in this review. In the areas of trauma, newborn intensive care, and transplantation, the provision of services has already been regionalized or is otherwise highly regulated by government or other administrative bodies. Other areas were generally limited to very few articles per condition.

We developed an initial set of inclusion criteria and tested it on studies of three conditions: cardiac surgery, carotid endarterectomy, and AIDS. After revision, the final inclusion criteria were:

1. Time: patient cohorts treated from 1980 forward.
2. Sample: community or population-based sample—case series or convenience samples were excluded.
3. Multiple publications from the same database excluded; only the most recent or most complete publication was included
4. Health outcome(s) must be assessed as the dependent variable(s).
5. Volume must be an independent variable.

We limited the review to studies of patients treated from 1980 to the present, because of the rapidity of changes in hospital care, available treatments, and surgical techniques. In our view, data from patient cohorts prior to 1980 would have questionable relevance to today's policy issues. In a few instances, we included studies if part of their patient sample included patients treated in 1978 or 1979, but most of the sample comprised patients from the 1980s. We excluded studies from single institutions, from voluntary registries, or other convenience samples because of the weak generalizability of such studies. We excluded a few studies in which the only dependent variable was a composite of deaths or long lengths of stay, because, formulated in this way, the dependent variable was not purely a health outcome. We also excluded a few studies in which the only dependent variable was a composite of death or complications, with the latter determined solely by secondary diagnosis codes in administrative databases. These studies were excluded because of the notorious unreliability of using such data to identify complications. In general, we excluded multiple publications from the same set of data, selecting only the most recent or complete, unless different publications reported substantially different analyses (e.g., one reported the relationship of hospital volume to outcome and another analyzed physician volume and outcome).

Three reviewers assessed the articles for inclusion or exclusion, with at least two reviewers independently examining each article and applying the criteria. Discrepancies in the application of the criteria were resolved by discussion between the reviewers. We developed an initial set of assessment criteria and tested it on the studies of CABG surgery, carotid endarterectomy, and AIDS which had passed the inclusion criteria. Our final criteria for quality assessment and the

scoring system were described earlier and are listed in Appendix A. The same pair of reviewers who assessed each article for inclusion or exclusion then independently evaluated each article and assigned quality scores. Discrepancies were resolved by discussion between the 2 reviewers.

Overall Results of Literature Review

Identification of Studies

The MEDLINE searches of subject headings resulted in 740 citations of all types of articles, including reviews, editorials, and letters. The MEDLINE searches of authors known to have published on the topic yielded 160 citations (including overlap with previous searches). Consultation with experts in the field yielded 136 citations (including overlap). The Cochrane Collaboration Database contained no relevant current or pending reviews.

We reviewed the titles and abstracts of the 850 citations. We were able to exclude 600 based on the title or the abstract. Four of these were because of a language other than English. 250 citations remained for possible inclusion into the study. Of these, 128 were about one of the eight chosen conditions. We found ten additional citations by examining the references of all included articles. We were unable to obtain the articles for one citation (Choti 1998). Thus, we reviewed a total of 137 articles for possible inclusion into our systematic review. We counted an article that studied more than one condition as more than one study. For example, we counted the article by Hannan published in 1999 as two studies because it examined CABG and AAA. The total number of studies was 162.

Exclusion of Studies

Of the 162 studies, 74 were excluded, and 88 were included. The most common reason for exclusion was because volume was not an independent variable (31 studies, 19%). The second most common reason for exclusion was a study sample that was not community or population-based (22 studies, 14%). This was particularly common among the studies of coronary angioplasty and cancer. These were primarily early studies of multiple surgical procedures that analyzed data from the 1970s. A few studies were excluded because the study outcomes were not health outcomes.

A summary of this information in Table 1 below.

Overall Results of Systematic Review

Overall, 88 studies were included in our systematic review. Two authors scored each study on the 10 item methodological quality scale described previously. Possible total quality scores range from 0 to 18. The actual distribution of quality scores ranged from 2 to 13 (See Figure 2). The mean total quality score was 7.8 ± 1.9. The median score was 8 with an interquartile range of 6 to 9.

TABLE 1 Results of Literature Search

Condition	No. of Studies Reviewed	No. of Studies Included	No. of Studies Excluded
Coronary Artery Bypass Graft	19	9	10
Pediatric Cardiac Surgery	3	3	0
Carotid Endarterectomy	24	18	6
Abdominal Aortic Aneurysm	16	12	4
Cancer Surgery	45	28	17
Coronary Angioplasty	20	9	11
Acute Myocardial Infarction	13	3	10
AIDS	9	6	3
Mulitple Procedures	13*	**	13
TOTALS	162	88	74

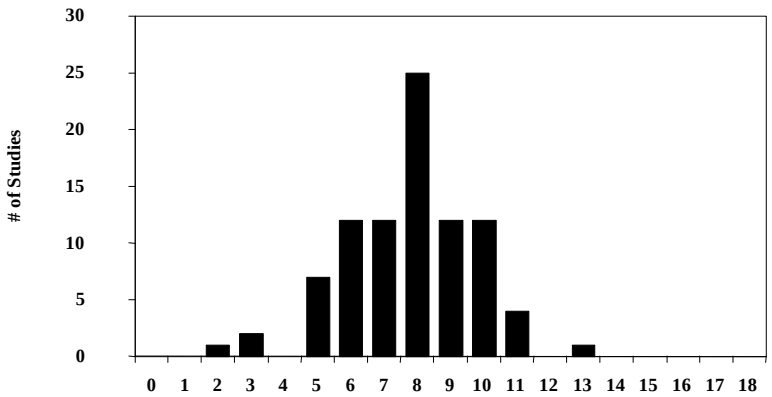
* We counted an article that studied more than one procedure as separate studies for each.

** The articles that studied more than one procedure that were included are reflected in the numbers for each procedure.

Since we excluded articles that analyzed data collected prior to 1980, the publication dates of those included in our review ranged from 1984 to 2000 (Our last MEDLINE search was March 2000). We found evidence of a dramatic increase in interest in this area as nearly half of all the studies we identified were published since 1998.

The methodological characteristics of the 88 articles we reviewed are as follows. Approximately, two-thirds (68%) of the studies were judged to be truly representative of the underlying population of interest (i.e, all patients undergoing the specified procedure). The remaining one-third of authors focused exclusively on the care of a certain category of patients by insurance status such as Medicare or Veterans Affairs patients. Nearly all of the studies were based on administrative data sources, so it was no surprise that 78% of them had sample sizes of 1000 cases or more—many had tens of thousands of cases. Approximately 80% of studies were based on

Figure 2 Distribution of Quality Scores of Volume/Outcomes Studies (N = 88)



samples that included 20 or more hospitals and 50 or more treating physicians. We used this measure as a rough indication that there was the potential for substantial variation at the two provider levels of interest. However, 9% of authors reported on samples with fewer than 20 hospitals and 50 physicians. Twelve percent of studies had an intermediate number of providers.

While we had no easy way of measuring the statistical power of each study, we did abstract the absolute number of primary outcomes (e.g., inpatient deaths) as a proxy indicator of the potential power of the study to detect modest associations or perform any significant multivariate modeling. It is worth noting that event rates are a function of both the overall sample size and the absolute rate of the adverse outcome of interest. Three-fourths of studies reported more than 100 adverse events upon which the volume-outcomes analyses were based. The rest of the studies had 20 to 100 events, except for one which had fewer than 20 (Wade 1996).

The primary outcome was mortality in 78% of the studies. This was most commonly inpatient death, though many reported 30 day death rates, and a few reported on 1, 3 or 5 year survival. The nineteen studies (22%) that examined more than just mortality were primarily studies of carotid endarterectomy that considered death or stroke as primary outcomes and investigations of coronary angioplasty that focused on death or emergency coronary artery bypass grafting together as major adverse events.

The vast majority of investigators examined provider volume as a categorical variable. Twelve percent dichotomized volume into low and high without any more detailed look for dose-response relationships. The remaining 88% reported at least three categories of volume, usually tertiles or quartiles, or numerous categories broken into even-numbered ranges. A few authors analyzed volume as a continuous variable.

Among the 88 studies we reviewed, there were 78 attempts to detect hospital level volume-outcome effects and 35 comparisons of physician level influences. The predominance of analyses of hospital volume effects is largely due to the fact that most of the studies were based on state or national hospital discharge databases that often lacked unique physician identifiers.

Most studies (72%) were primarily focused on examining either the effects of hospital or physician volume on outcomes. Another 10% of studies reported both hospital and physician levels effects, but separately. Sixteen studies (18%) performed some type of multivariate or stratified analysis designed to try to tease apart the independent or synergistic contributions of physician and hospital volume to outcomes.

We paid particularly close attention to the level of sophistication of the risk adjustment techniques used by investigators (if any), because failure to account for imbalances in patient case mix among providers can seriously confound the relationship between volume and outcome. Fifteen percent of studies performed no risk adjustment at all. The majority of studies (58%) relied exclusively on risk adjustment models based on administrative data, most commonly sociodemographics and ICD-9 discharge diagnoses. The limitations of these risk adjustment techniques are beyond the scope of our discussion but have been reviewed elsewhere (Iezzoni 1997). Only 24 studies (27%) used any clinical data in their risk adjustment models. Among this group, only six studies reported risk adjustment models that were robustly discriminating (model C statistic ≥ 0.75) and calibrated (by Hosmer-Lemeshow or other testing).

There were two other higher level methodological issues that few studies seemed to address. Only 2 of 88 studies (2%) measured the appropriateness of patient selection—both were studies of carotid endarterectomy where this important issue has been recognized for a decade (Brook 1990a, Karp 1998). Similarly, only 4 of 88 authors (5%) made any attempt to adjust for the use of any key processes of care. Two studies of breast cancer adjusted for use of adjuvant therapy or

type of surgery (Sainsbury 1995, Roohan 1998). Thiemann and colleagues' (1999) study of myocardial infarction adjusted for the use of key medications (e.g., aspirin and beta-blockers) and revascularization. Malenka et al (1999) adjusted for use of heparin or intra-aortic balloon pump in coronary angioplasty cases.

Finally, while we found marked heterogeneity in the methodologic quality of studies for each publication year we reviewed, we were interested in more formally assessing whether the overall quality of published research was improving over time. The results of this preliminary analysis are displayed in Figure 3. We did find a significant, though modest, increase in quality scores over time ($R = 0.26$, $p = .014$). Studies with more recent publication dates received higher scores than earlier ones, suggesting that the level of analytic sophistication in this field is improving.

Volume and Outcome in Coronary Artery Bypass Graft (CABG) Surgery:

Nineteen articles treating various aspects of the subject of volume and outcome in CABG surgery were identified by our various search strategies. Ten failed to meet our inclusion criteria. Two reported data from hospitals that were not community or population based samples. (Zelen 1991, Clark 1996) One was a review article and presented no primary data. (Snowden 1995) Three dealt with various issues concerning regionalization of CABG surgical services but 2 presented no data on outcomes (Anderson 1989, Rohrer 1997), and one had no data on the relationship between volume and outcome. (Roos 1989) Two were excluded because either mortality by hospital was not evaluated (Banta 1990), or because the relationship of neither hospital nor physician-specific volume was examined with respect to outcomes. (Tu 1996). One was excluded

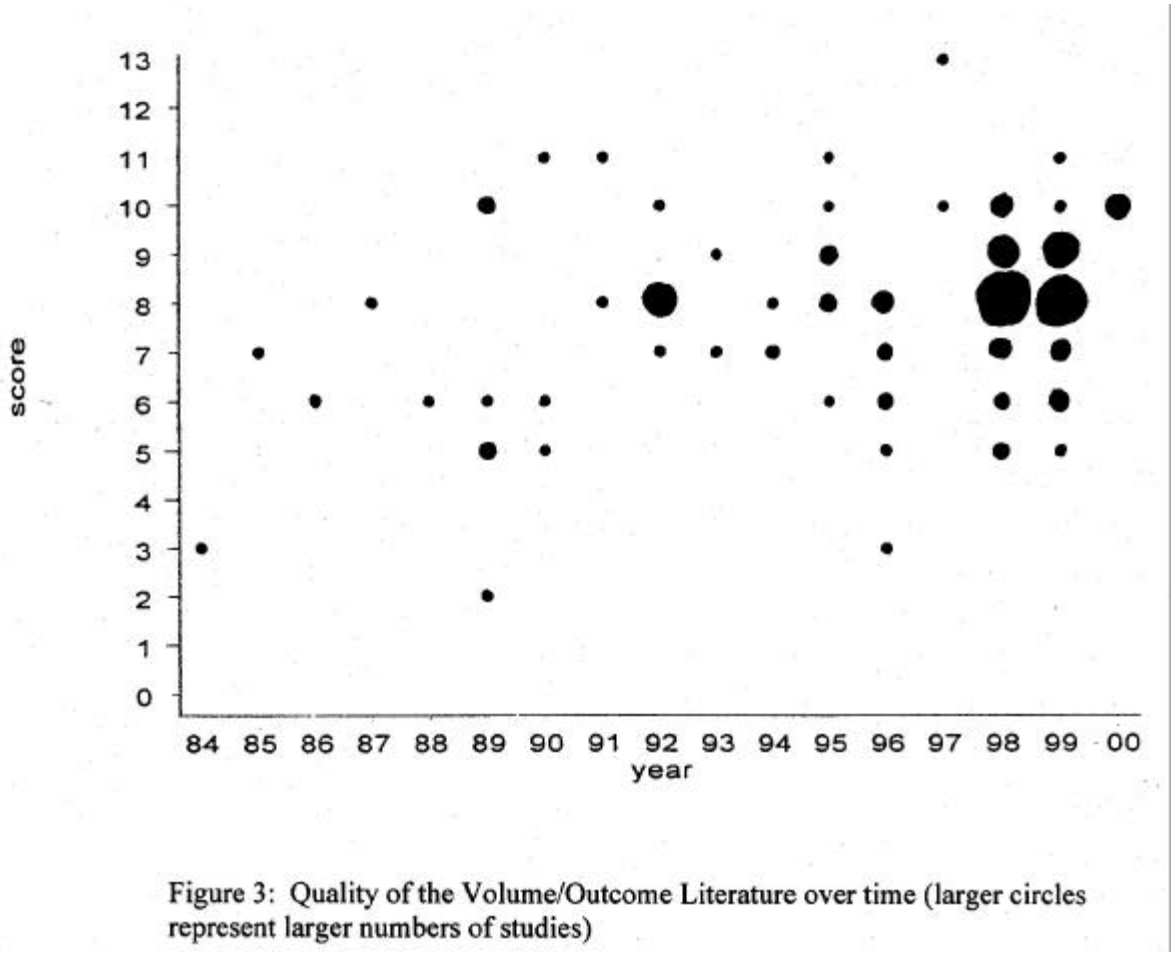


Figure 3: Quality of the Volume/Outcome Literature over time (larger circles represent larger numbers of studies)

because health outcome was not the dependent variable; that study used a composite variable of inpatient death plus long length of stay. (Hughes 1987). And one study was excluded because, although risk-adjusted mortality data and hospital volume data were presented together, their association was not explored. (Hannan 1994)

Of the 9 articles meeting our inclusion criteria, overall quality scores ranged from 7 to 11, with a median of 8. These studies and their main characteristics are listed in Appendix C. Two studies received a score of 7, four obtained scores of 8, and one each received scores of 9, 10, and 11. The three highest scoring studies achieved these ratings either by examining the joint effects of hospital and surgeon volume (n = 2) or by using clinical data to risk-adjust (n = 2), or both (n = 1). No studies examined any outcomes other than death, and none assessed appropriateness of patient

Figure 3: Quality of the Volume/Outcome Literature over time (larger circles represent larger numbers of studies) selection or clinical processes of care. All 9 studies employed risk-adjustment methods in analyzing outcomes; 7 used administrative data and 2 used clinical data. All 9 studies had large sample sizes (n > 5,000); 7 studies examined data from more than 10,000 cases, and 4 had sample sizes greater than 50,000.

Seven of the 9 studies demonstrated a statistically significant association between mortality and surgeon or hospital volume, or both. The earliest study was limited by its use of Medicare data to assess the relationship of death (within 60 days of surgery) to hospital volume of CABG operations on Medicare patients in 1979–1980. (Riley 1985) Because older people received this operation much less often 20 years ago, Medicare hospital volumes were very low by today's standards; 76% of cases were performed in hospitals with volumes less than 100. Showstack and colleagues used administrative data from California hospitals in 1983 to demonstrate that risk-adjusted inpatient death rates were higher at hospitals performing 100 CABG operations or fewer (5.2%) compared with those performing at least 350 such surgeries (3.1%). Grumbach and colleagues demonstrated similar findings in their study of CABG surgery in California, New York, and Canada (Ontario, Manitoba, and British Columbia). (Grumbach 1995)

Hannan and colleagues have published the only investigations to examine the impact of both surgeon and hospital volume on mortality, finding significant inverse relationships for both. They first examined administrative data from New York State in 1986 and found that high-volume surgeons (>116 cases) had lower risk-adjusted inpatient mortality across all strata of hospital volume than lower-volume surgeons. High-volume surgeons at high-volume hospitals (>650 cases) experienced a risk-adjusted mortality of 3.2% compared with low-volume surgeons at low-volume hospitals (<224 cases), who demonstrated a risk-adjusted mortality of 5.7%. (Hannan 1989) In the most detailed and rigorous assessment of these relationships, these authors used prospectively collected clinical data to risk-adjust 1989 inpatient mortality in New York hospitals. (Hannan 1991) They again found statistically and clinically significant associations between mortality and both surgeon and hospital volume. Table 2 summarizes these findings.

Two studies employed several years' of data to examine changes in the volume outcome relationship over time in CABG surgery. Farley used administrative data from 62 hospitals participating in the Hospital Cost and Utilization Project from 1980 to 1987. They examined the relationship between inpatient mortality and hospital volume both cross-sectionally and longitudinally. (Farley 1992) They found that volume at low-mortality hospitals increased over time and interpreted the findings as evidence for the "selective-referral hypothesis."

TABLE 2 Risk-Adjusted Inpatient Mortality (%) Following CABG Surgery: New York State in 1989

	Surgeon Volume				
Hospital Volume	<55	55–89	90–259	260	Total
<200	14.07	6.67	6.42	3.33	7.25
200–889	9.01	5.72	3.97	2.88	4.32
890	5.33	3.39	3.11	2.18	2.85
Total	8.14	5.56	3.61	2.43	3.68

Hannan and colleagues used clinical data to examine changes in surgeon volume and their relationships to risk-adjusted inpatient mortality in New York from 1989 to 1992. (Hannan 1995) They demonstrated that risk-adjusted mortality for low-volume surgeons (50 cases per year) fell 60%—from 7.94% to 3.20%—during the study period. This decline was greater than that observed for high-volume surgeons (34%) and for all cases (43%). In a closer analysis of their data, however, they depicted an important dimension of complexity not previously described. In each of the 4 years studied, the composition of the low-volume surgeon group changed markedly. Although the number of low-volume surgeons was similar in each of the 4 years (between 33 and 39), only 10 surgeons remained in the low-volume group in all 4 years. In each year, some new surgeons entered practice, some became low-volume surgeons transiently, and some stopped doing CABG surgery altogether. The 10 persistently low-volume surgeons had the highest risk-adjusted mortality rate in the final study year of any identified subgroup. Their mortality rate in 1992 was 8.96%, compared to all low-volume surgeons in that year (3.2%), low-volume surgeons who were new to New York that year (2.07%), low-volume surgeons who had previously had higher volumes (1.82%), or the statewide average (2.45%).

Two studies found no association between volume and outcome. Shroyer and colleagues examined data from 44 Veterans Affairs Hospitals performing CABG surgery between 1987 and 1992 and found no relationship between hospital volume and death within 30 days of surgery. (Shroyer 1996) The range of volume among the hospitals studied was limited. No hospital performed more than 400 CABG operations, and only 3 performed between 300 and 400. The lack of hospitals in high or very high volume strata (as characterized in other research) precluded the ability of this study to detect an association between outcome and a more complete range of volume. In contrast, Sollano et al examined the experience in New York State between 1990 and 1995 and found no relationship between inpatient death and hospital volume. (Sollano 1999) In this study, too, the range of volume was limited, but in an opposite direction compared with the Shroyer study. Only 1 of 31 hospitals averaged fewer than 100 CABG operations during the study period, 1 averaged between 100 and 199 procedures, and 4 averaged between 200 and 299 operations. In this sample, 18 hospitals (58%) averaged more than 400 cases per year, a volume not reached by any of the hospitals in the Shroyer study. It seems likely that the Sollano investigation lacked sufficient hospital representation in the low-volume range to adequately assess the association between volume and outcome in CABG surgery.

Volume and Outcome in Pediatric Cardiac Surgery

Three studies met our inclusion criteria for this review. No studies were excluded. Appendix D summarizes the characteristics, findings, and quality scores of the 3 studies. Quality scores

were modest (6, 8, and 9). None of the studies examined appropriateness or processes of care or studied an outcome other than inpatient mortality. None of the studies included large numbers of hospitals, because pediatric cardiac surgery is concentrated in a very small subset of hospitals, much smaller than the number performing adult cardiac surgery. All 3 studies took a similar analytic approach to examining how mortality was related to volume of surgery for congenital heart disease in children. In contrast to many adult surgical procedures, the frequency of occurrence of specific pediatric cardiac surgical procedures (e.g., repair of a ventricular septal defect or correction of tetralogy of Fallot) is too low to support comparisons of volume across different hospitals. All 3 of the research teams assessing the relationship between volume and outcome in this area divided into 4 complexity categories all of the surgical procedures performed to correct congenital heart disease in children, aiming to construct categories with relatively homogeneous mortality risk. The volume variable in these analyses is then the total number of all such procedures, with the complexity category entering a multivariate model as an independent risk-adjustment variable.

Jenkins and colleagues used administrative data from California (for 1988) and Massachusetts (for 1989) and were the first to create complexity categories. (Jenkins 1995) They constructed their categories using clinical judgment to sort ICD9 procedure codes into 4 groups. When they applied the resulting classification scheme to their data set, they did demonstrate increasing average mortality rates across the 4 categories: 4.8%, 6.4%, 10.7%, and 15.9%. In their multivariate model, adjusting for demographic characteristics and complexity, volume was strongly related to inpatient mortality. Risk-adjusted inpatient mortality was 18.5% for hospitals at which fewer than 10 cases were performed, 7.9% for those in the range of 10–100, 8.2% for those performing between 101 and 300 cases, and 3.0% for hospitals at which more than 300 of these operations were done. Although the 95% confidence intervals for these estimates overlapped considerably for the lower 3 volume categories, the confidence interval for the estimate of the highest volume category did not overlap with any of the others. The authors did not present data on model calibration or discrimination, but used generalized estimating equation procedures to control for intra-hospital correlations.

Sollano and colleagues used the complexity categories created by Jenkins et al. in their analysis of administrative data from New York State (1990–1995). (Sollano 1999) They used standard logistic regression to assess the relationship between hospital volume and inpatient mortality, adjusting for procedure complexity, and found a significant odds ratio of 0.944 for hospital volume (in hundreds of cases), suggesting a decrease of 5.6% in the odds of dying for every additional 100 cases in volume. The average overall mortality was 7.3%. The model's C statistic was 0.617; calibration was good. Further analysis by age group showed that the relationship of volume to outcome was especially strong for neonates, age 30 days, (OR = 0.64) and for infants, age 31 days–1 year, (OR = 0.72) but not statistically significant for children over 1 year of age.

Hannan et al. took a different analytic approach to constructing complexity categories and had access to prospectively collected clinical data to evaluate volume and inpatient mortality in New York State from 1992 to 1995. (Hannan 1998a) These researchers ordered the cardiac surgical procedures by their actual mortality rates in their data set. Expert clinicians then grouped the procedures into clinically sensible subgroups and repeated the process until 4 categories were identified. This procedure resulted in a classification system with more widely separated average mortality rates than that created by Jenkins et al. Death rates by category were 1.4%, 4.5%, 11.0%, and 20.1%. These categories were employed as independent variables in a multivariate

TABLE 3 Risk-Adjusted Inpatient Mortality (%) Following Surgery for Congenital Heart Disease in New York State (1992–1995)

	Surgeon Volume		
Hospital Volume	<75	75	Total
<100	8.94*	7.45	8.26*
100	8.47	5.45**	5.95**
Total	8.77*	5.90**	6.75

*Higher than state average ($p < .05$).
**Lower than state average ($p < .05$).

logistic regression, which also controlled for age and comorbidities (as identified in the clinical database). The resulting model demonstrated excellent calibration and discrimination (C = 0.818). The researchers used the model to explore the relationship between inhospital death and volume for both hospitals and surgeons. Table 3 shows the findings. Both the hospital and surgeon effects were statistically significant. The authors chose the volume categories that demonstrated the greatest differentiation between low and high volume providers. In general, they noted that higher volumes were associated with lower risk-adjusted mortality rates across all categories of procedure volumes.

In each of the 4 complexity categories, both high-volume hospitals and high-volume surgeons demonstrated lower risk-adjusted mortality than their lower-volume counterparts.

Volume and Outcome in Carotid Endarterectomy

We were able to identify 19 studies of carotid endarterectomy that met our inclusion criteria (See Appendix E). Six additional studies initially identified by our search algorithm were excluded (2 because there was no explicit volume/outcome analyses (Tu 1998, Hsia 1992) and 4 because they were not population-based (Hertzer 1984, Rubin 1998, Friedman 1988, Slavish 1989).

In general, there are two types of studies in the literature. The most common type is based on a state or national hospital discharge database and reports on ten of thousands of carotid endarterectomies, but has limited clinical information. These studies have a large number of cases, but tend to have very limited if any risk-adjustment, and infrequently have both surgeon and hospital information. They are also limited to mortality as the only believable outcome, since post-operative stroke can not be reliably ascertained by administrative data. This is because stroke can be a pre-operative diagnosis as well as a post-operative complication, and ICD-9 discharge codes do not distinguish between the two. The other type of study is much smaller (usually 500 to 1500 patients), but has more detailed clinical information from medical chart review that facilitates measurement of stroke and death rates (usually over 30 days) and allows more clinically grounded risk adjustment models.

There was considerable heterogeneity in the overall quality scores ranging from 3 to 10 (out of a possible 18 point scale). The median score was 6 (inter-quartertile range 5–8). Only 9 of the 19 studies performed any type of risk adjustment. Among studies that did attempt to adjust for patient risk factors, five relied on administrative data, and four had clinical data. All four that collected clinical data received a quality score of 8 or greater. Only two studies (Hannan 1998b,

Khuri 1999) presented moderately robust risk-adjustment models with C statistics of 0.715 and 0.72, respectively. Only two studies attempted to adjust the analyses for appropriateness of the procedure, and none adjusted for processes of care.

The unit of analysis was also quite variable. Four studies looked solely at surgeon volume, six examined only hospital volume, and nine examined both (3 of which examined them together). Nearly all studies reflect carotid endarterectomy practice from the 1980s and mid-1990s. Half (10) of the studies considered inpatient mortality to be the primary outcome. The other half (9) focused on 30 day death rates, some with combined death, stroke, and myocardial infarction rates. Because of the well-documented difficulties of using administrative data to differentiate pre-operative versus post-operative strokes, we only considered the 4 studies that had medical chart review data to have valid measures of perioperative stroke or myocardial infarction.

While a few studies considered procedure volume as a continuous variable, most investigators divided annualized caseload rates into two to five categories and compared the outcomes of the lowest versus the highest-volume providers. There was ten-fold variation in the definition of low and high-volume providers among the studies. The definition of low volume surgeon ranged from ≤1 per year to <30 per year; and the definition of low-volume hospital ranged from ≤6 per year to ≤100 per year. This heterogeneity makes formal meta-analysis virtually impossible.

Of the 13 studies that analyzed surgeon volume, nine found some positive relationship between surgeon volume and outcomes. Of the four highest rated studies (scores of ≥9), three found that the low volume surgeons (<5 per year to <10 per year had significantly worse risk-adjusted outcomes compared with highest-volume surgeons. The one higher-quality study by Cebul and colleagues (1998) that found no effect of physician caseload had a relatively small sample size (n = 678) and absolute number of adverse events (n = 32). The authors themselves calculated that they had less than 10% statistical power to find a significant relationship between observed differences in high compared with low-volume physicians. However, their study did have the statistical power to find a significant relationship between hospital volume and outcome.

Of the 15 studies that analyzed hospital level effects, seven found a significant relationship (and one a borderline trend p = .08) with clinical outcomes and institutional volume. Of the three higher rated studies (scores of ≥ 9) that measured volume effects, one (Katonen 1998) conducted in Finland found no hospital effect, and the two in the U.S. (Cebul 1998 and Hannan 1998b) did find significantly worse outcomes for hospitals with ≤ 62 and ≤ 100 carotid endareterectomies per year.

Among the eight studies that examined both surgeon and hospital volume, only two performed any risk adjustment. Cebul and colleagues (1998) found that only hospital volume was significantly associated with risk-adjusted 30 day stroke and death rates, not surgeon volume. Hannan et al. (1998b) which included data on 28,207 operations over a six year period in New York found both hospital and surgeon volume to be independently associated with risk-adjusted inpatient mortality rates, with the surgeon effect being larger than the hospital effect. Hannan et al. also found a significant interaction between hospital and surgeon volume as shown in Table 4. Low volume surgeons operating at low volume hospitals had the lowest risk-adjusted mortality rates (1.96%), low volume surgeons at high volume hospitals and high volume surgeons at low volume hospitals had identical, intermediate mortality rates (1.18%), and high volume surgeons

TABLE 4 Risk-Adjusted Inpatient Mortality (%) Following Carotid Endarterectomy in New York State (1990–1995)

	Hospital Volume		
Physician Volume	1–100	101	Total
1–4	1.96*	1.18	1.89*
5	1.18	0.94**	1.11**
Total	1.28*	0.94**	1.19

*Higher than state average ($p < .05$).
**Lower than state average ($p < .05$).

operating at high volume hospitals had the lowest death rates (0.94%). It is also worth noting that in New York State, there were very few operations performed by low volume surgeons in high volume hospitals, so estimates of mortality rates in this subgroup had very wide 95% confidence intervals.

Volume and Outcome in Abdominal Aortic Aneurysm Surgery

Sixteen studies on abdominal aortic aneurysm (AAA) repair were reviewed. Four were excluded. Two were excluded because volume was not an independent variable (Evans 1999, Adam 1999). One study (Veith 1991) was excluded because it was based on the same data as another study (Hannan 1992). Another study was excluded because its data was from the 1970s (Pilcher 1980). A summary of the articles on AAA is provided in Appendix F.

Twelve studies met the inclusion criteria. Overall quality scores ranged from 5 to 10, with a median of 8. These moderate scores reflect the frequent use of administrative databases for risk adjustment and the exclusive use of mortality as the outcome. The unit of analysis was the hospital in seven studies, the hospital and the physician (analyzed separately) in three studies (Pronovost 1999, Dardik 1998, Rutledge 1996), and the hospital and the physician analyzed together in two studies (Hannan 1989, Hannan 1992). No study measured the appropriateness of patient selection.

The definition of a low-volume hospital was heterogeneous, making comparisons among studies difficult. It ranged from less than four to less than 50 unruptured AAAs per year. The definition of a low-volume surgeon was more homogeneous, ranging from less than three to less than five unruptured AAAs per year.

A somewhat unique characteristic of many of the AAA studies was an attempt to measure the volume of something other than the index condition, and to determine the association of this volume with the outcomes of that condition (i.e., the issue of “volume of what”). These studies asked whether experience with vascular surgical procedures other than the one under study might be associated with the outcomes of the studied procedure. Three studies examined the relationship between unruptured AAA volume and ruptured AAA outcomes (Dardik 1998, Rutledge 1996, Amundsen 1990). One study looked at total AAA volume and outcomes for ruptured AAA and unruptured AAA (Manheim 1998). Three studies looked at the relationship between the volume of operations other than AAAs and AAA outcomes (Khuri 1999, Amundsen 1990, Hannan

1992). Khuri and Amundsen measured the volume of all peripheral vascular operations, and Hannan looked at the volume of any aorta operation—i.e. not just aortic aneurysm repairs.

All but the Amundsen study performed some type of risk adjustment, and the Khuri study of Veterans Affairs hospitals was the only one to use clinical data for risk adjustment. The Khuri study was the only study in this entire review to perform hierarchical linear modeling. Its C statistic for AAA was 0.75.

No study measured clinical processes of care. Pronovost et al (1999) came the closest, measuring the effect of intensive care unit characteristics on outcomes. This included physician staffing, (e.g. daily rounds by an ICU physician), nurse staffing (e.g. nurse-patient ratios), and some in practice policies (e.g. use of a critical pathway, morbidity and mortality reviews, routine extubation in the operating room). All of these variables, except for routine extubation in the operating room, are structural characteristics or general policies and not actual clinical processes of care measured at the patient level. Routine extubation in the operating room was analyzed only in terms of its effect on length of stay, and not on any clinical outcome.

Every study looked at only one outcome – mortality. Khuri measured 30-day mortality, and all other studies measured inpatient death. Complications of AAA repair such as bleeding, renal failure, lower extremity ischemia, ischemic colitis, myocardial infarction, and respiratory failure were not measured.

All but one study (Khuri 1999) found a significant relationship between volume and outcomes. Six of the seven studies that looked exclusively at hospital volume found higher mortality among low-volume hospitals. In a study of unruptured AAAs in New York State, Sollano et al found that an increment of 100 cases per year was associated with a 22% decrease in the odds of death (Sollano 1999). Similarly, in a study of Ontario hospitals, Wen et al found that an increase in hospital volume of 10 unruptured cases per year was associated with a 6% decrease in the odds of death from unruptured AAA (Wen 1996). They found no such relationship for ruptured AAAs.

Both Dardik and Rutledge looked at outcomes for ruptured AAAs. Dardik et al found that hospital volume (unruptured or ruptured) was not related to outcome of ruptured AAAs, but that surgeon volume of ruptured cases was (Dardik 1998). Similarly, Rutledge et al found only a non-significant trend between hospital volume (unruptured or ruptured) and outcome, and a significant relationship for surgeon volume (Rutledge 1996).

Two studies received a quality score of 10 (Hannan 1989, Hannan 1992a). These studies had large numbers of patients, physicians, and hospitals, and they examined surgeon volume, hospital volume, as well as the surgeon and hospital volume together. Hannan et al evaluated all AAAs performed in New York State in 1986 (Hannan 1989). Hospital volume was related to outcome only when the surgeon was a high-volume surgeon ($\geq$4 cases per year). When the surgeon was high-volume, high-volume hospitals (>36 cases per year) had a risk-adjusted mortality of 11%, and low-volume hospitals had a risk-adjusted mortality of 19%. Surgeon volume was related to outcome, regardless of the hospital volume. Overall, high-volume surgeons had a risk-adjusted mortality of 15%, and low-volume surgeons (<4 cases per year) had a risk-adjusted mortality of 20%. The effect of surgeon volume was particularly strong within high-volume hospitals. High-volume surgeons operating in high-volume hospitals had a risk-adjusted mortality rate of 11%, whereas low-volume surgeons in high-volume hospitals had a risk-adjusted mortality of 20%. The study also found that for high-volume surgeons, mortality continues to decrease as hospital volume increases over the measured range of hospital volumes.

In another study of New York State, Hannan et al (1992a) examined the relationship of hospital and surgeon volume to outcomes of unruptured and ruptured AAA over a three year period (1985 to 1987). They found that for unruptured AAAs, hospital volume was related to outcome, but that surgeon volume was not. By contrast, for ruptured AAAs, surgeon volume, but not hospital volume, was related to outcome. They repeated the analysis using the volume of all aorta operations and found the same results.

In order to try to evaluate the "selective referral" hypothesis, they identified groups of hospitals and groups of physicians with the lowest and highest mortality rates in 1982–1984, and then followed their changes in volume during the period 1985–1987. Although all groups of surgeons experienced some increase in volume, low-mortality surgeons had the greatest increase in volume (523 operations to 751 operations, or 44% increase). To evaluate the "practice makes perfect" hypothesis, they looked at the outcomes for those low-volume surgeons in 1982–1984 who became high-volume surgeons in 1985–1987. There were 22 low-volume surgeons who increased their volumes to 14 or more cases per year. The decrease in their risk-adjusted mortality rate was negligible (6.8% to 6.2%). Six of those 22 surgeons, however experienced an increase in volume to 22 or more cases per year, and the drop in their mortality rate was larger (5.8% to 2.5%), although still not statistically significant because of the small number of surgeons. The authors concluded that some evidence for selective referral as well as the practice makes perfect hypotheses exists.

The only negative study was Khuri and colleagues' (1999) evaluation of all AAAs performed in Veterans Affairs (VA) hospitals from 1991 to 1997. In this study, 107 hospitals performed 3,767 procedures. Thirty-day mortality was measured, and clinical data was used to build a fairly robust risk-adjustment model. Annual hospital volume ranged from 0 to 32, with a mean of 6.9. The study found no relationship between hospital volume and outcome for AAAs or for any of the seven other procedures it examined. The findings of this study may not be generalizable, because it included only VA hospitals and patients. The authors point out that many VA surgeons operate at other institutions as well, so the volume of the VA institutions may not in fact reflect the experience of their staffs.

Volume and Outcome in Cancer Surgery

We examined a total of 38 studies on cancer. All of the eight studies of medical treatment of cancer were excluded because none of them looked at volume as an independent variable. Of the 30 studies of surgical treatment, 10 were excluded. The most common reason for exclusion was a sample that was not community or population-based (7 studies). Two studies did not evaluate volume as an independent variable (Gordon 1998, Whittle 1998). One paper was a review article, not primary research (Steele 1996).

Thus, 20 papers, all about cancer surgery, were included in the systematic review. Three of these studies looked at more than one procedure (Hannan 2000, Gordon 1999, Begg 1998). To analyze these articles, we examined the data for each procedure separately. In total, 11 papers studied pancreatic resection, five studied colorectal resection, three studied esophagectomy, three studied lung resection, and two studied breast surgery (summarized in Appendix G). The three articles that looked at other cancer procedures are summarized in a table in Appendix G called "Cancer Miscellaneous."

We did not include other papers that studied these operations for benign as well as malignant disease, with the exception of Gordon (1999). We included Gordon (1999) because it studied

pancreaticoduodenectomy and esophagectomy, both of which are rarely performed for benign disease.

Pancreatic Resection

Eleven studies evaluated pancreatic resection. The quality scores varied greatly, ranging from 3 to 10, with a median of 7. The study with the lowest quality score had a small sample that was not representative of the entire population and did not perform any risk adjustment (Wade 1996). The study with the highest quality score had a large, representative sample, and it examined physician volume, hospital volume, and the interaction between the two (Lieberman 1995).

The unit of analysis was the hospital for all studies, except for two that looked at both surgeon and hospital volume (Lieberman 1995, Sosa 1998a). No study evaluated appropriateness of patient selection. The definition of low hospital volume ranged from less than 1 to less than 9 procedures per year. Begg et al defined volume as the annual volume of procedures done on Medicare patients. Two studies of Maryland had only one high-volume hospital (Gordon 1995, Gordon 1999). In Lieberman and colleagues' study of New York State, two hospitals were high-volume, and four surgeons were high-volume. The two analyses of surgeon and hospital volume interaction were limited by the fact that most of the high-volume surgeons practiced only in high-volume hospitals.

No study effectively addressed the question of "volume of what." Gordon et al studied the association between the total volume of 6 "complex gastrointestinal" procedures (total colectomy, esophagectomy, total gastrectomy, hepatic lobectomy, biliary tract anastomosis, and pancreaticoduodenectomy) and individual procedure mortality. They did not also study, however, the association between individual procedure volumes and mortality (Godron 1999). No study evaluated the appropriateness of patient selection.

Risk adjustment was based almost exclusively on administrative data. Only Begg et al used some clinical data (cancer staging from the Survival, Epidemiology, and End Results database). None of the studies examined clinical processes. Inpatient death was the primary outcome of interest. Three studies looked at death beyond the inpatient stay (Simunovic 1999, Birkmeyer 1999a, Birkmeyer 1999b), and one measured rates of complications, specifically infection and hemorrhage (Glasgow 1996). Other complications such as pancreatic or biliary leak, gastric dysmotility, pneumonia, and other outcomes such as recurrence and quality of life were not examined.

Of the nine studies that looked at hospital volume only, all but one (Wade 1996) found a significant relationship between volume and outcomes. The highest quality score of 8 was achieved by a study of 1705 pancreatectomies at 298 hospitals in California from 1990 to 1994 (Glasgow 1996). In this study, the risk-adjusted mortality at high-volume hospitals (>50 cases per year) was 3.5%, compared to 14% at low-volume hospitals ($\leq$5 cases per year).

Lieberman et al (1995) analyzed both physician and hospital volumes; 1,972 procedures were performed by 748 surgeons in 184 hospitals in New York State from 1984 to 1991. In separate analyses of surgeon volume and hospital volume, high-volume surgeons ($\geq$41 cases per year) had lower risk-adjusted mortality rates than low-volume surgeons (<9 cases per year)—6% versus 13%, and high-volume hospitals (>8 cases per year) had lower risk-adjusted mortality rates than low-volume hospitals (<10 cases per year)—5% vs. 19%. When surgeon volume and hospital volume were analyzed together, however, only hospital volume was significant.

Sosa et al (1998a) analyzed both physician and hospital volumes for 1,236 procedures by 373 surgeons at 48 hospitals in Maryland. They found that the relative risk of death at low-volume hospitals (<5 cases per year) was 19 times that at high-volume hospitals (>20 cases per year). Analyzing physician and hospital volume together, they found hospital volume to be significant regardless of physician volume.

Although the studies on pancreatic resection had a great deal of methodological heterogeneity, they suggested that outcomes were related to provider volume and to hospital volume in particular. The magnitude of this volume effect was relatively large compared to most of the other procedures we studied. This is a function of both the high absolute mortality rate for pancreatic cancer as well as a very strong volume and outcome relationship. The number needed to be treated by a high volume provider to prevent one inpatient death attributable to low volume was only 10 to 15 for most higher-quality studies.

Esophagectomy

The three studies of esophagectomy had low quality scores (6, 6, and 8). The two lower-scoring studies had relatively small sample sizes – 518 patients in one (Gordon 1999) and 503 patients in another (Begg 1998). The unit of analysis was the hospital in all three studies. The definition of low volume was relatively similar across studies, ranging from less than 6 to less than 10 procedures per year. Begg et al measured volume of Medicare cases only. All studies performed some risk adjustment, and only one utilized clinical data (Begg 1998). No study evaluated clinical processes such as operative approach (abdominal versus thoracoabdominal) and method of reconstruction.

The only outcome evaluated was inpatient mortality. No study examined long-term survival, recurrence, or quality of life. Complications such as anastomotic leak, respiratory failure, pneumonia, and digestive dysfunction were not measured.

All three studies found large differences in mortality between low-volume and high-volume hospitals. Gordon and colleagues found that the relative risk of death at a low-volume hospital was 3.8 times that at a very-high-volume hospital, although there was only one institution in this latter category (Gordon 1999). Begg et al found that the risk-adjusted mortality at high-volume hospitals was 3.4%, compared to 17.3% at low-volume hospitals. Patti et al. (1998) found similar mortality rates—6% at high-volume hospitals and 17% at low-volume hospitals. This study had the highest quality score of 8, in part because of its large size. Overall, the magnitude of the volume and outcome relationship for esphagectomy was striking. The number needed to treat by a high volume provider to prevent one inpatient death attributable to low volume was seven to nine patients.

Breast Cancer Surgery

The two studies of breast cancer surgery had relatively high quality scores (10 and 11) because they had large numbers of patients, surgeons, hospitals, and adverse events, and because they utilized clinical data from cancer registries in their risk adjustment models. The unit of analysis was the hospital in one study (Roohan 1998) and the surgeon in the other (Sainsbury 1995). Neither study looked at the appropriateness of patient selection. Roohan et al defined "very low" hospital volume as fewer than 10 cases per year. Sainsbury et al defined low surgeon volume as fewer than 30 cases per year. Sainsbury et al attempted to include extent of disease

and tumor grade in their risk-adjustment model, though this information was missing for 50% of patients.

The two studies were noteworthy for their measurement of clinical processes. Roohan et al included the type of operation (mastectomy or breast-conserving surgery) as an independent variable in the multivariate analysis. Sainsbury et al included the percentage of patients treated by mastectomy (versus local excision), chemotherapy, hormone therapy, radiation therapy, or surgery alone for each surgeon. These two studies were unique in that they both selected a long-term outcome (5-year survival) as their dependent variable. Neither study measured other outcomes such as recurrence, complications of surgery, or complications of adjuvant therapy.

Roohan et al looked at 47,890 cases of breast cancer surgery performed in 266 hospitals in New York State from 1984 to 1989. In a multivariate regression model, they found volume to be related to 5-year mortality, with a clear "dose-response" relationship. The increased risk of death was 19% in moderate-volume versus high-volume hospitals, 30% in low-volume versus high-volume hospitals, and 60% in very-low-volume versus high-volume hospitals. The authors conjectured that since breast surgery has negligible operative and inpatient mortality, the volume-outcome relationship might be caused by higher-volume hospitals providing more effective adjuvant treatment.

Sainsbury et al studied 12,861 cases of breast cancer surgery performed by 180 surgeons in the Yorkshire Regional Health Authority area from 1979 to 1988. Risk adjustment included age, extent of disease, tumor grade, socioeconomic status, date of treatment, and type of therapy (surgery, radiation, chemotherapy, hormone therapy, surgery alone). They found that the risk of death was significantly lower for patients of high-volume surgeons (greater than 29 cases per year) compared to low-volume surgeons (fewer than 10 cases per year). There was no difference in survival between moderate-volume (10 to 29 cases per year) and low-volume surgeons. The volume effect was slightly smaller after risk adjustment (risk ratio of 0.86 versus 0.82 before adjustment). Variation among surgeons in use of mastectomy, radiation, chemotherapy, hormone therapy, and surgery alone accounted for 8% of the variation in survival. Surgeon volume and use of chemotherapy accounted for 20 to 25% of the variation in survival.

Lung Resection

The quality scores of the three studies of lung resection were relatively high (8, 8, and 10). The numbers of patients, physicians, hospitals, and adverse events were all high. The unit of analysis was the hospital in two studies (Begg 1998, Romano 1992) and both hospital and physician in one study (Hannan 2000). No study evaluated the appropriateness of patient selection. The three studies looked at different types of lung resection—lobectomies (Hannan 2000), pneumonectomies (Begg 1998), and all resections (Romano 1992).

The definitions of low hospital volume were heterogeneous, ranging from less than 6 to less than 38 procedures per year. Risk adjustment was based on administrative data in two of the studies (Hannan 2000, Romano 1992) and clinical data in one (Begg 1998). No study looked at clinical processes of care. The outcome of interest was inpatient death in all three studies. Complications such as bronchopleural fistula, respiratory failure, and pneumonia were not measured. In addition, no study evaluated other outcomes such as long-term survival, recurrence, or quality of life.

In the study with the highest quality score of 10, Hannan et al (2000) looked at 6,954 lobectomies by 373 surgeons at 178 hospitals. The risk-adjusted mortality rate at low-volume hospitals

(≥37 cases per year) was 1.65% higher than at high-volume hospitals (≥169 cases per year). There was no difference between medium-volume and high-volume hospitals. The vast majority of hospitals were low-volume (133 hospitals). Only 4 hospitals were high-volume. No significant relationship between surgeon volume and outcome was found.

Begg and colleagues examined 1,375 pneumonectomies performed on Medicare patients at 313 hospitals in the U.S. They utilized clinical data for risk adjustment. No difference in outcomes existed between high-volume and low-volume hospitals. Romano and colleagues found 40% lower risk of death after pneumonectomy at high-volume hospitals compared to low-volume hospitals. They also found a similar volume-outcome relationship for lesser resections.

Colorectal Resection

The five studies of colorectal cancer resection had quality scores ranging from 7 to 10, with a median of 9. The studies were very heterogeneous. Three studies evaluated resections of all types of colorectal cancer (Hannan 2000, Harmon 1999, Parry 1999), one looked at total colectomy for benign and malignant disease (Gordon 1999), and one looked at resections for rectal cancer (Porter 1998). The unit of analysis was the hospital in 1 study (Gordon 1999), the physician in one study (Porter 1998), and both hospitals and physicians in 3 (Hannan 2000, Harmon 1999, Parry 1999).

The definition of low volume was variable, even among the three studies that looked at volume of all colorectal resections. Among these three studies, the definition of low surgeon volume ranged from less than 6 to less than 12 procedures per year. The definition of low hospital volume ranged from less than 40 to less than 84 per year. Gordon et al looked at the relationship between 6 complex gastrointestinal procedures including total colectomy and the outcomes of total colectomy.

All studies performed risk adjustment, and two studies (Porter 1998, Parry 1999) used clinical data. Two studies examined clinical processes, but neither incorporated the processes into their risk adjustment model. Parry et al measured whether or not an abdominoperineal resection was performed, use of ultrasound or CT scan, and operating "after hours." Porter et al looked at the type of operation (low anterior resection versus abdominoperineal resection) and the use of adjuvant therapy. The outcome studied was primarily inpatient mortality. One study (Parry 1999) measured local recurrence rates as well as disease-specific survival. No study measured complications such as anastomotic leak, intra-abdominal abscess, wound infection, or genitourinary dysfunction.

Three of the four studies that assessed hospital volume did not find a significant relationship to outcomes. Harmon et al studied all resections in Maryland and found a trend toward lower mortality at high-volume hospitals, but this was not significant (odds ratio 0.78, $p < 0.10$). Parry et al studied all resections in the Northwestern United Kingdom and found no relationship between volume and outcomes. Gordon et al found no relationship between volume of complex gastrointestinal surgery and outcome of total colectomy. The only study to find a significant relationship for hospital volume found that the risk-adjusted mortality rate at low-volume hospitals was 1.9% higher than at high-volume hospitals (Hannan 2000).

Of the four studies that measured physician volume, three found a significant volume-outcome relationship. Only Parry et al found no relationship between physician volume and outcomes. Porter et al found that patients of low-volume surgeons had worse disease-specific survival than patients of high-volume surgeons (hazard ratio = 1.40) and a higher risk of local recur-

rence (hazard ratio = 1.80). High-volume surgeons were more likely to perform a low anterior resection as might be expected. They were no more likely, however, to use adjuvant therapy than low-volume surgeons were.

Three studies analyzed physician volume and hospital volume together. The physician effect found by Hannan et al disappeared when hospital volume was controlled for in the analysis. Harmon et al found that surgeon volume was related to volume regardless of hospital volume.

The studies of volume and outcome in colorectal surgery do not uniformly find a significant relationship. The magnitude of the volume effect on mortality is relatively modest—an absolute difference in inpatient mortality of 1% to 2% corresponding to a number needed to treat of 50–100.

Summary

The twenty studies of cancer surgery suggest that a significant relationship between volume and outcomes does exist. The largest differences between low and high-volume providers were found for the most complicated operations in rare cancers—pancreatectomy and esophagectomy. For colorectal resection and lung resection, two operations for more common cancers, the relationship between volume and outcome is not as clear.

The common methodologic issues for these studies point to a need for more clinical data. Information about the type of tumor and cancer stage would be highly desirable, particularly in studies that look at long-term survival. An examination of the different clinical processes being employed and how they vary with provider volume, might elucidate the differences in outcomes. For example, the use of adjuvant therapies is particularly important but has not been well-studied with respect to volume. The roles of other providers besides the surgeon have also not been examined. Particularly when long-term survival is being evaluated, characteristics of other providers who care for the patient years after surgery, such as the medical oncologist and radiation oncologist would be relevant. More appropriate referral to these providers or better coordination of the many elements of cancer care, such as diagnostic testing, adjuvant therapy, and follow-up surveillance, may underly the hospital volume effects that have been found.

It is worth noting that the literature on volume and outcomes in cancer has disproportionately focused on rare operations for rare cancers. For the most common cancer operations – breast cancer surgery, colon resection, and lung resection, we found 10 studies that met our inclusion criteria. By contrast, the most rare operations – esophagectomy and pancreatectomy had 13 publications. In addition, we found no studies of medical treatment of cancers.

Volume and Outcome in Coronary Angioplasty

Nine studies on the relationship between provider angioplasty volume and outcomes met inclusion for our systematic review (See Appendix H.) Eleven other studies were identified by our search algorithm but excluded for the following reasons: two did not have volume and outcome data (Gray 1999, de Belder 1999), two were based on data from the early 1970s, and seven were not population-based samples—either small convience samples (Ellis 1997, Kastrati 1998, Klein 1997, Shook 1996) or nested within a randomized controlled trial (Gilchrist 1999) or trial registry (Grassman 1997, Kimmel 1995).

All of the studies were based on tens of thousands of patients. Seven of the nine were derived from large statewide or national hospital discharge databases. Overall quality scores ranged from 7 to 14, with a median score of 8 (IQR, 8–9). All of these articles included some attempt at

risk adjustment, though most of them (6) did this using administrative data from hospital discharge abstracts. The exceptions were Hannan (1997a), Malenka (1999) and McGrath (1998) which used databases that had more detailed clinical data that could be used for more rigorous risk adjustment. The study by Hannan and colleagues (1997a) using the New York State Coronary Angioplasty Reporting System reported the most robust risk adjustment model with a C statistic of 0.89. No studies adjusted for appropriateness, though Malenka (1999) adjusted for clinical indication. Also of note, only one study, Malenka et al. (1999) adjusted for important processes of care.

The unit of analysis among the the studies also varied. Five studies examined hospital volume alone, two looked at physician volume alone, and two analyzed both. All studies reflect practice from the late 1980s through the mid-1990s. None of the articles included data on angioplasty performed after 1996, so these data do not reflect the use of newer antiplatelet therapies as well as other more recent innovations in technique.

Compared with several of the other conditions which we reviewed, the literature on angioplasty was somewhat more homogenous. All studies were very large databases of tens of thousands of patients, included attempts at risk adjustment, and examined the outcomes of inpatient death and emergency (or same stay) coronary artery bypass grafting (CABG). Emergency CABG is an important outcome because it usually is performed when angioplasty is unable to open the target clogged coronary artery or when the procedure results in a dangerous complication such as coronary artery dissection. A few studies included other additional outcomes such as 30 day survival (Jollis 1994 and Jollis 1997) and radiographically defined successful angioplasty (patency; McGrath 1998, Malenka 1999).

Nearly all investigators compared outcomes of providers across a broad spectrum of volume categories, most commonly tertiles or deciles of caseload. The definition of low volume physician ranged from <25 per year to ≤85 per year. It is worth noting that the American College of Cardiology/American Heart Association guidelines recommend a minimum physician volume of 75 procedures per year. This is particularly significant in interpreting the finding of the two articles from the Northern New England Cardiovascular Disease Study Group (NNEDSG; Malenka 1999, McGrath 1998), which used as their lowest-volume physician volume thresholds < 84/yr or ≤ 85/yr which are often used to define “high-volume” providers in policy statements and other analyses. Relative to the rest of the country, the NNEDSG was really a study of high-volume, very high-volume and extremely-high-volume providers. This may explain why both investigators were unable to find any relationship between procedure volume and mortality. The definition of low-volume hospital ranged from <50 per year to <400 per year. The ACC/AHA guidelines recommend a minimum hospital caseload of 200 angioplasties per year.

Of the four studies that analyzed the effects of physician volume, only one found a significant association between procedure volume and mortality alone (Hannan 1997a), and one found a trend towards such a relationship (Malenka 1999, $p = .086$). Three studies found that lower angioplasty volume correlated with higher rates of emergency CABG. Emergency CABG rates average about 3% to 5% compared to mean inpatient mortality rates of about 1.0 to 1.5%, so studies that measured rates of both types of major angioplasty complications likely had greater statistical power to detect modest influence of volume on outcomes.

Of the seven articles that examined hospital volume, three found an inverse relationship between volume and mortality alone, though four did not. However, all seven found that lower volume was associated with higher rates of emergency CABG alone as well as the combined endpoint of inpatient death or emergency CABG. The two studies that examined both physician

and hospital volume influences (Hannan 1997a, Jollis 1997), both reported that physician and hospital procedure volume were independent predictors of outcomes in multivariate analyses. That is, adjusting for physician volume, hospital volume was still important, and vice versa. While neither investigator found any significant formal interactions between physician and hospital volume, both observed the lowest risk-adjusted event rates among the highest volume cardiologists performing angioplasty in the highest volume hospitals.

Three studies stand-out for their methodological sophistication. The highest rated study by Hannan and colleagues (1997a) examined all 62,670 procedures performed in New York in 1991 to 1994 using the New York State Department of Health Coronary Angioplasty Reporting System database. This dataset contains details on several important pre-procedural clinical variables identified by the New York State Cardiac Advisory Committee as necessary for rigorous risk adjustment including location, percent stenosis, and morphology of the diseased coronary vessels. This level of granularity of information may help explain how Hannan and colleagues were able to develop a robust risk adjustment model with excellent model fit and discrimination (C statistic = 0.89 and Hosmer-Lemeshow test not significant). They found that patients undergoing angioplasty by cardiologists with annual volumes < 75 per year had significantly higher risked adjusted rates of inpatient mortality and emergency CABG than the state as a whole (death: 1.03% vs. 0.90%; CABG: 3.93% vs. 3.43%). Hospital volume was also important. Hospitals that performed <400 angioplasties per year had higher risk adjusted rates of inpatient death and same stay CABG compared to the statewide average (death: 1.12% vs. 0.90%; CABG: 4.19% vs. 3.43%). They also found that both cardiologist and institutional caseload made independent contributions to explaining the variance in clinical outcomes. They found no formal interactions between physician and hospital volume.

Two additional points are worth comment. Hannan and colleagues (1997a) compared the outcomes of the lowest volume categories to the statewide average, not to the highest volume categories. For all of the comparisons, there did not appear to be a smooth linear relationship of decreasing event rates for increasing volume, in part, because the highest-volume categories tended to have higher event rates than the intermediate volume categories. In addition, the relative difference in outcomes was fairly modest resulting in an approximate 13% difference in results by cardiologist volume and 19% difference by hospital volume. The absolute difference was even smaller, yielding a number needed to treat to prevent one death attributable to low volume of 454 for cardiologist caseload and 769 for hospital caseload.

Jollis and colleagues (1997) used the national Health Care Financing Administration hospital discharge abstracts to compare the outcomes of 119,886 Medicare beneficiaries undergoing angioplasty in 1992. Their volume categories are based on numbers of Medicare patients undergoing angioplasty, who are thought to account for only one-third to one-half of angioplasty patients, depending on the physician or institution. They relied on administrative data for risk adjustment. The authors reported that higher physician volume was associated with lower risk adjusted rates of the combined endpoint of in-hospital death or CABG, even after controlling for hospital volume. Similarly, they found that higher hospital volume was associated with lower risk adjusted rates of both in-hospital death and the combined outcome of death or CABG, even after controlling for cardiologist volume. However, the actual numerical estimates of the risk-adjusted analyses were not reported in the paper.

The unadjusted numbers were reported and are as follows. Low-volume cardiologists (<25 per year) had higher unadjusted rates of the combined endpoint of inpatient death or CABG compared to those with high-volume (>50 per year)—6.0% versus 4.7%. Low-volume hospitals

(<100 per year) had higher unadjusted death or CABG rates than those with high-volume (>200 per year)—6.5% versus 4.9%. The magnitude of these relative differences range from 22% to 25% and correspond to numbers needed to treat to prevent one death or CABG attributable to low provider volume of 77 and 62, respectively. However, they found no significant differences in unadjusted or adjusted inpatient or 30 day mortality rates by volume categories.

Finally, the study by Malenka et al. (1999) is noteworthy because it was one of the few studies in our entire systematic review that collected information on key processes of care and also adjusted for these variables in their multivariate modeling, in this case, the use of heparin, coronary stenting, or intra-aortic balloon pump.

Volume and Outcome in Acute Myocardial Infarction

While there is an enormous health services research literature on acute myocardial infarction, we were only able to identify three studies that met our inclusion (See Appendix I.) Eight other studies were identified by our search algorithm but excluded because they did not explicitly evaluate volume and outcome relationships. Most of these articles focused on the effects of physician specialty or hospital type (teaching vs. non-teaching) on clinical outcomes. One study (Barbash 1994), did explicitly analyze the effects of hospital volume, but was excluded because it was based on a thrombolytic trial database, which was considered a non-population based sample.

All three studies that did meet our inclusion criteria utilized very large hospital discharge databases that contained data on ten of thousands of patients. Thiemann and colleagues (1999) received the highest quality rating of the group because they utilized a robust, clinically-based risk adjustment model and also adjusted for important processes of care known to influence outcomes following myocardial infarction, such as certain medications (aspirin and beta blockers) and revascularization procedures.

Thiemann (1999) used the Health Care Financing Administration Cooperative Cardiovascular Project database that contains clinical and administrative information on a nationwide sample of Medicare patients with acute myocardial infarction during an eight month period during 1994–1995. In addition, to the standard data, the CCP contains information on important processes of care known to influence outcomes, such as the use of aspirin, beta-blockers, ACE inhibitors, thrombolytic therapy, and revascularization procedures. They compared 30 day and one year survival rates according to quartiles of hospital volume defined as <1.4 MI patients/wk, 1.4–2.5 per week, 2.6–4.4 per week and >4.4 per week. Unadjusted 30 day mortality rates in the four groups ranging from lowest to highest volume were: 16.7%, 15.3%, 14.7%, and 14.4%, respectively.

Using a detailed clinical risk adjustment model (C = 0.79), they found that compared to the lowest hospital volume category, the hazard ratio for 30 day survival was 1.17 for the highest volume quartile, 1.07 for the second highest quartile and 1.05 for the third quartile. Overall, a decrease of 5.5 Medicare patients per week was associated with a hazard ratio for 30 day death of 1.10 and a long term hazard for 1 year mortality of 1.05. While the risk of death was highest among the lowest-volume hospitals (<1.4 per week), hospital volume remained a continuous predictor even among the two highest volume quartiles.

The authors also stratified by the specialty of the admitting physician, and found that the association between low hospital volume and higher mortality rates was consistently found among patients treated by cardiology, internal medicine, or family practice physicians. Interestingly,

they found no relationship between individual physician volume and survival in models with and without physician specialty.

The Theimann (1999) article is most noteworthy for being one of the few studies in our systematic review that examined the independent influence of key processes of care on the volume-outcomes relationship. For example, they reported that adjusting for differences in processes of care such as use of aspirin, thrombolytic agents, beta-blockers, ACE inhibitors, and revascularization procedures accounted for about one third of the survival benefit attributed to high volume hospitals. This indicates that different patterns of use of specific treatments or procedures can explain part, though not all, of the variations seen in outcomes of patients treated in high versus low-volume hospitals.

Casale and colleagues (1998) used administrative database from Pennsylvania to analyze 30,715 admissions for acute myocardial infarction in 1993. The Pennsylvania database also contained clinical data on patient severity using the MedisGroup classification (now called the Atlas System). This clinical data enabled the researchers to produce one of the more robust risk adjustment models identified by our review (C statistic = 0.881). They divided hospitals into low volume providers, those below the median number of myocardial infarction patients (≤ 12 per year) and high volume, those above the median (>12 per year). Patients treated by physicians a high volume of cases had 11% lower odds of risk-adjusted inpatient death (OR = 0.89).

Of note, the primary purpose of the study by Casale and colleagues was to examine the effects of physician specialty on outcomes in myocardial infarction. They reported that patients treated by cardiologists had lower in-hospital mortality rates than those treated by primary care physicians (OR = 0.87). However, much of this effect was probably mediated through physician volume effects as 85% of cardiologists were high volume providers compared to 31% of primary care physicians. In addition, high volume remained an independent predictor even after controlling for physician specialty. This study serves as a harbinger that many of the specialist versus generalist debates in the outcomes literature may be seriously confounded by volume effects.

The study by Farley et al. (1992) took a somewhat different approach by focusing on the potential effect of changes in provider volume over time. This is in comparison to the vast majority of articles we reviewed that evaluate volume and outcome associations in a cross-sectional manner. Farley and colleagues used longitudinal data from the Hospital Cost and Utilization Project to analyze patterns of myocardial infarction care in a national sample of 600 non-Federal hospitals over the period of 1980 to 1987. One of the things these authors did differently from most others was to logarithmically transform volume to adjust for skewness (the asymmetric distribution of many low volume hospitals and only a few very-high-volume ones with extremely large caseloads). They then considered the natural log of volume [ln(volume)] as a continuous variable. In analyses that used administrative data (including the Disease Staging severity system using ICD-9 discharge diagnosis) for risk adjustment, they reported that a 10% increase in caseload corresponded to a 2.2% absolute decrease in inpatient mortality rates.

Using a set of sophisticated statistical models, they concluded that the strength of the relationship between hospital volume and inpatient mortality increases when one moves from analyzing cross sectional data (as most investigators have done) to longitudinal or “panel” data. Examining data within a given hospital over time increases the likelihood of adequately controlling for hospital-specific factors that could influence both volume and outcomes and thus bias estimates of the direct relationship between volume and outcomes. They also conducted a set of regressions using instrumental variables that they assert shows no selective referral effect for the conditions of myocardial infarction, inguinal hernia repair and neonatal respiratory distress syn-

drome. Their study also suggests that other factors such as "nursing intensity," which are often unmeasured, may be associated with inpatient mortality rates.

Volume and Outcome in AIDS Treatment

Nine research articles on AIDS were reviewed. Three failed to meet inclusion criteria for the systematic review. One was excluded because the same data on outcomes and volume were reported in an earlier article by the same researchers. (Bennett 1990). One was excluded because the patient sample was not community or population based; it reported data from a single institution. (Wachter 1986) The third study was excluded because volume was not an independent variable in its analyses; it investigated the experiences of generalists and specialists as the "predominant providers" for patients with AIDS. (Turner 1994)

Six studies met the inclusion criteria and were analyzed in the review. (See Appendix J for summary overview.) Overall quality scores were low: 3 studies received scores of 7, 2 received scores of 6, and 1 received a 2. Principal common factors underlying these low scores is that all of the investigations assessed death as the only outcome and none evaluated the effect of appropriateness or processes of care. Five studies examined the relationship between outcome and hospital volume alone and one examined the association between survival and physician volume. None examined both hospital and physician experience together. All used at least administrative data to risk-adjust rates of death, and 2 used clinical data but none presented data documenting good discrimination or calibration. All six demonstrated statistically significant inverse associations between hospital or physician volume and death rates. Patient sampling, the formulation of the volume variable, and analytic presentations differed widely among the 6 studies, precluding meta-analysis and rendering generalizations hazardous.

Bennett and colleagues first reported this relationship in a study of administrative data from 15 California metropolitan area hospitals. (Bennett 1989) They studied inpatient mortality for patients admitted with AIDS and their first episode of Pneumocystis carinii pneumonia (PCP) in 1986–1987. They compared outcomes at hospitals with high "familiarity" with AIDS (30 admissions of patients with HIV infection per 10,000 total admissions) to those with less familiarity (<30 AIDS-related admissions per 10,000). Unadjusted death rates were 33% at low AIDS-experience hospitals and 12% at high experience hospitals. In their multivariate model (using administrative data to risk-adjust), the odds of a patient dying in a hospital with low-AIDS familiarity were about 3.6 greater than in a hospital with high AIDS familiarity.

Using a similar approach with administrative data from 73 New York City hospitals, Bennett and colleagues also studied inpatient deaths for patients with AIDS and first episodes of PCP in 1987. The volume variable in this study was the number of cases of AIDS-related PCP treated in the study year. Hospitals were classified as low-volume (<160 cases) or high volume (160 cases). High-volume hospitals exhibited an unadjusted death rate of 21% compared with 30% at low-volume hospitals. Logistic regression using the administrative data for risk-adjustment demonstrated statistically significant odds ratios of 0.64 (for hospital volume of 160–239 cases) and 0.67 (for 240 cases), both compared with hospitals with fewer than 80 cases.

Stone et al. used clinical data from a state AIDS registry to risk-adjust both inpatient and 30-day death rates for patients with AIDS admitted to Massachusetts hospitals in 1987–1988. (Stone 1992) These investigators chose to define volume as the number of discharges of patients with AIDS per 10,000 discharges at each hospital. The 20% of hospitals with the greatest volume were designated "high-experience" (43 AIDS discharges per 10,000), the remainder as "low-experience"(<43 AIDS discharges per 10,000). The unadjusted relative risk of dying in a low-

experience hospital was 2.16 times that in a high-experience institution (19.0% vs. 9.8%). The unadjusted relative risk of dying within 30 days of admission was similar (1.93). When adjusted by logistic regression for a variety of demographic and severity of illness variables derived from the AIDS registry, the relative risk of inpatient death remained significant (RR = 2.92 at low versus high-experience hospitals) as did the relative risk of death within 30 days of admission (RR = 2.51).

Turner and Ball used administrative data from a nationally-representative sample of 258 hospitals (from the Hospital Cost and Utilization Project) in 1986–1987 to examine the relationship between death during hospitalization for patients with any manifestation of AIDS and the total number of AIDS discharges in each hospital in each study year. (Turner 1992). In a logistic regression model controlling for demographics and comorbid conditions, an increase in a hospital's volume of AIDS cases of 100 was associated with a decrease of 4.3% in the odds of an inpatient death.

In the only study of hospital care to examine more than 2 years of data, Hogg et al used administrative data to assess inpatient mortality for all patients with AIDS admitted to all Canadian hospitals (n = 513) from 1987 through 1994. In a logistic regression model controlling for age, year of admission, and clinical stage, the average number of AIDS admissions per year was significantly associated with inpatient mortality. Compared with hospitals experiencing fewer than 1 AIDS admission per year, hospitals with 1–9 demonstrated an odds ratio of 0.79; 10–99, OR = 0.77; and those with 100, OR = 0.64.

In the sole study in which the dependent volume variable was physician experience, Kitahata and colleagues found a significant relationship between survival from the time of diagnosis and physician experience with AIDS. (Kitahata 1996) In their sample of 403 patients with AIDS treated by physicians at Group Health Cooperative of Puget Sound, they found a decreased relative risk of death for patients treated by physicians whose experience consisted of more than 5 AIDS patients, compared with those with low experience (1 patient). Using clinical data to adjust for severity of illness and CD4 counts, the relative risk of death was 0.57 for patients of high-experience physicians. The median survival from diagnosis was 14 months for low-experience physicians compared with 26 months for their more experienced colleagues.

Despite their methodologic weaknesses, all six included studies showed statistically significant and, in some cases, clinically impressive associations between hospital volume and mortality (usually inpatient mortality). Whether a publication bias exists that acts against the submission or publication of studies demonstrating no association between volume and outcome in this disease is unknown. None of the 5 studies examining hospital experience assessed physician experience or measured specific processes of care. These studies did little to illuminate the underlying reasons for the associations they documented. No study reported data from the recent era of AIDS treatment, characterized by the emergence of protease inhibitors and an increasing array of complex multidrug regimens. The role of the hospital has diminished greatly in the care of patients with AIDS. How generalizable the findings of these studies are to present-day AIDS care is unclear.

Conclusions

Several conclusions emerge from this systematic review of the research since 1980 on the association between volume and outcome in health care. There can be little doubt that for a wide variety of medical conditions and surgical procedures, patients treated at higher volume hospitals or by higher volume physicians experience on average lower mortality rates than those treated by

low-volume hospitals and physicians. The 88 studies we reviewed contained 113 different assessments of the relationship between hospital or physician volume and outcomes; 53 examined hospital volume only, 10 focused on physician volume only, and 25 assessed both. Counting the 25 studies that evaluated both as separate assessments, 77% of all the assessments in the studies in our review found statistically significant associations between either hospital or physician volume and health outcomes. In all cases the association was in the direction of higher volume being associated with better outcomes. No study documented a statistically significant association between higher volume and poorer outcomes. The remaining 23% of evaluations found results that were not statistically significant. This conclusion is strengthened by the further observation that all 16 studies with the highest quality scores (10 or greater) demonstrated statistically significant associations.

The majority (60%) of studies investigated hospital volume effects only, a smaller proportion examined only physician volume (11%), 10% assessed hospital and physician volume but as separate independent variables, and 18% analyzed their relative contributions taken together. Across all conditions and procedures, an almost identical proportion of studies reported statistically significant, positive associations for better outcomes and hospital volume (78%) as for physician volume (74%). Of the 16 studies reporting the independent effects of both physician and hospital volume, 12 (75%) found statistically significant independent effects for both, 3 (19%) found a significant effect for hospital but not physician volume, and 1 (6%) found a significant physician volume association but none for hospitals. A majority (63%) of these 16 studies examined data from one state (New York).

One important cautionary note must be appended to our degree of confidence in the conclusion that the association between volume and outcome is firmly established. We cannot exclude the existence of a significant publication bias that would diminish the likelihood of studies failing to document such associations appearing in the research literature. Although it is possible that such a bias might operate at the level of review and editorial decisions at individual journals, it is also possible, and perhaps more likely, that such a bias could operate at the level of individual investigators, who choose not to pursue or report analyses that fail to demonstrate significant associations. Such decisions could occur because of loss of interest in the research topic or fear a bias on the part of journal editors against studies reporting negative findings. In addition, when faced with data about a number of procedures or conditions, researchers may choose to report only the positive findings.

Risk-adjustment methods varied in their sophistication, with the most rigorous using prospectively collected clinical data to adjust outcomes for severity of illness and comorbidities. Many of these risk-adjustment models achieved high levels of statistical performance in discrimination and calibration. In the vast majority of these studies, statistically significant associations between outcomes and hospital or physician volume were documented. It is thus most unlikely that differences in severity of illness or distribution of comorbid illnesses explain the association between volume and outcome.

The strength and magnitude of the statistically significant relationships described in these studies varied greatly. Many of the included studies analyzed data from tens of thousands of cases, having sufficient sample sizes to detect statistical significance even in small differences in outcomes. In absolute terms, significant differences between high and low volume hospitals ranged from about 1 death (or fewer) per 100 patients treated in PTCA and carotid endarterectomy to 5–10 per 100 in some studies of CABG, pediatric cardiac surgery, and AAA surgery.

The greatest differences were seen in some studies of complex cancer surgery (e.g., operations on the pancreas or esophagus), where differences as high as 10–15 deaths per 100 were reported.

Overall, the methodological rigor of the published research was modest, with only 18% receiving quality scores of 10 or higher; 75% of these received a score of 10, which is 56% of the maximum score of 18, and only 1 study received the highest score (13), which is 72% of the maximum. Thus, while the research documents the fact that high volume is often associated with better health outcomes, many questions about the nature of the relationship, its causes, and significance remain unanswered. For example, the studies in our review examined a wide variety of volume categories for the same procedures, typically comparing 2 or 3 volume strata or assessing the linear association between volume and outcome. Thus, we know little about whether a volume threshold exists (or more than one) above which outcomes are more favorable but do not continue to fall with further increases in volume. We do not know how these relationships vary among procedures or conditions or how different they are when considering physician volume.

Further, the correlation between volume and outcome, even if statistically significant, is true only on average (Hannan 1999a). Some low-volume hospitals will have outcomes comparable to the best of the high-volume hospitals and some high-volume hospitals will have poor outcomes. Moreover, in any given time period, some portion of the variability among hospitals (or physicians) will be due to chance and not to systematic differences in quality. This variability is not independent of volume. Lower volume hospitals or physicians will demonstrate greater variability, because proportions based on smaller sample sizes will generate greater chance variability compared with larger ones.

This effect is even more pronounced when the event representing the numerator of the proportion is uncommon. Figure 4 illustrates this phenomenon by plotting the hospital mortality rates for all 5705 hospitals treating one or more Medicare beneficiaries with an acute myocardial infarction in 1984 (Chassin 1989). The plotted mortality rates were adjusted for differences in age, sex, and race. As volume increases, the variation among hospitals in mortality rates diminishes dramatically. Further study of these data revealed that 62% of the variability among all hospitals in death rates was due to chance, even after using clinical data to adjust for differences among hospitals in severity, comorbidity, and processes of care (Park 1990). Extant research does not address how much of the observed outcome differences between high and low volume hospitals might be due to chance. This deficit in our knowledge is due in part to the lack of studies using longitudinal data to assess hospital or physician performance.

Another consequence of the paucity of longitudinal studies is the lack of insight provided into the question of the shape and duration of the learning curve, either for new physicians learning standard procedures or treatments or for experienced physicians learning new procedures. A related, rarely-asked question is what level of volume is required to maintain proficiency once acquired. The questions about the learning curve and maintenance of proficiency are equally relevant (and equally unanswered) for hospitals and other important members of the clinical team in addition to the treating physician.

In only 4 procedures were outcomes other than death evaluated—stroke in carotid endarterectomy, emergency CABG surgery in PTCA, local recurrence in colon resection, and bleeding and infection in pancreas surgery. Most of these studies showed the same association with these non-fatal outcomes as with mortality—that they occurred less frequently when hospitals or physicians had more experience. Many other vital health outcomes have not been investigated (e.g., physical or psychological functioning and other local surgical complications).

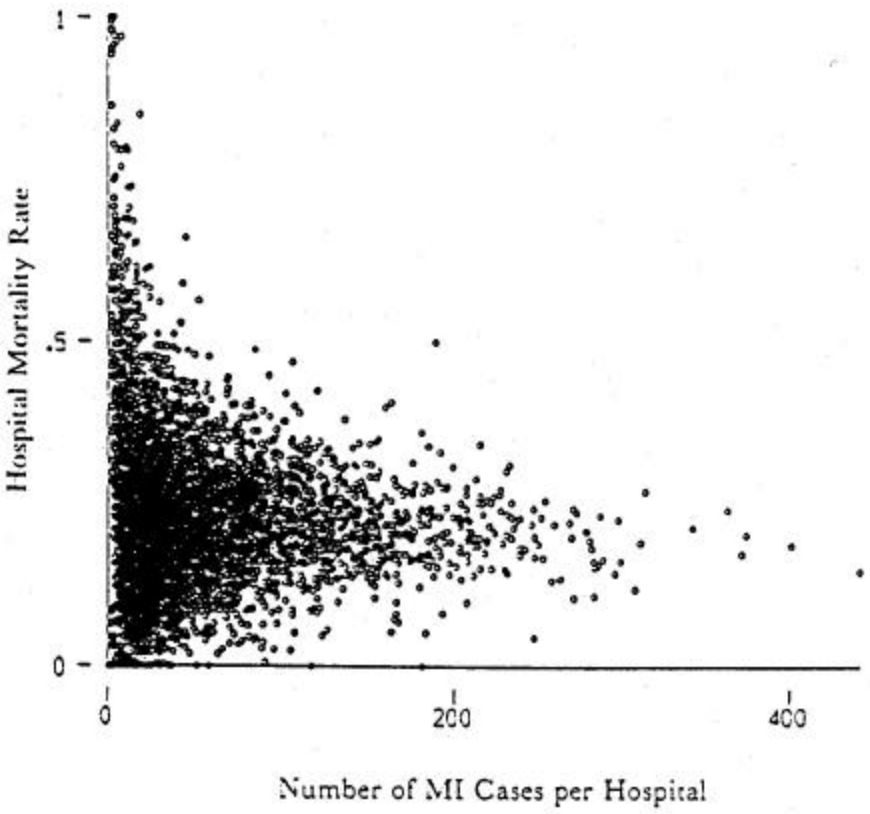


Figure 4: Variations in Acute Myocardial Infarction Mortality: 1984 (From Chassin 1989)

Current research did not address the extent to which (if any) the differences between high and low volume providers may be associated with differences in appropriateness of patient selection. Nor does current knowledge permit conclusions to be drawn about what unit of volume is most highly correlated with outcome. For example, a surgeon's proficiency with colon resection may be affected not only by his or her experience with that procedure, but also by experience with small bowel resection, because some very similar surgical tasks are involved in both kinds of procedures. The few studies that have addressed this question have reached variable conclusions and suffered from a lack of methodological rigor. Thus, the "volume of what" question remains unanswered.

This research has also failed to illuminate what underlying causes produce the association between volume and outcome. It is obvious that volume itself cannot directly produce good (or bad) outcomes. Other features of patient care, correlated with volume, are the most likely explanatory factors. Our conceptual model (Figure 1) suggests that specific clinical processes of care that bear a direct relationship to outcome may be performed more often in high-volume hospitals (or by high-volume physicians). The converse may also play a role (that harmful processes are performed less often by high-volume providers). The research literature we reviewed provided scant opportunities to test this hypothesis, as specific clinical processes were rarely studied. In one important study on elderly acute MI patients, the differential use of proven-effective treatments (thrombolytics, aspirin, beta blockers, and ACE inhibitors) explained about one-third of the survival benefit attributable to high-volume hospitals (Thiemann 1999). This

study documented that patients at high-volume hospitals with lower mortality rates were more likely to receive effective pharmaceutical treatments than those treated at low-volume hospitals.

We believe that future research directed at uncovering differences in processes of care between high- and low-volume providers carries the best chance of explaining why outcome differences exist between these groups. This research should include investigations of the relationship of the experience of other important members of teams of caregivers (e.g., consultants, intensivists, operating room teams). Organizational practices should also be studied, following on work done in intensive care (Shortell 1994, Pronovost 1999) and surgical services. (Young 1998) In addition, research of this type could help establish the effectiveness of specific processes, particularly in surgery, where such evidence is lacking today. Advancing knowledge in this direction could also provide a blueprint for improvement by identifying precisely what poor-performing providers (whether low-volume or not) need to do to improve.

Policy Implications

Is there a role for public policy to play in the application of this body of knowledge? Our discussion of policy implications is predicated on the assumption that health care policy in this area may have one or both of the following objectives: to inform consumers and clinicians about important information on outcomes and hospital and physician characteristics that may be important in decision making, or to improve patient outcomes.

We see four alternative policy directions that might address these objectives:

1. Data on volume might be made publicly available.
2. State governments might take action to limit the number of hospitals permitted to perform certain procedures.
3. Private purchasers might require or provide incentives for managed care plans or employees to steer patients to high volume providers.
4. Government (or other state or regional entities) might use data on volume and outcomes to improve performance.

A program of public availability of data on hospital or physician volume of services would have the advantage of access to reliable and reasonably timely data (at least on hospital volume) from many, if not most state hospital discharge abstract data bases. In the absence of reliable and valid data on outcomes, data on volume may be of interest to consumers and useful to referring physicians. Such a program would face the difficulty of explaining in understandable terms the complex relationships between volume and outcome that research has delineated, being careful to clarify what is known and what is not. For example, it will be important to explain the limitations in our knowledge about whether volume thresholds exist for different procedures or conditions, for hospitals, or for physicians. Updating data on a regular basis to reflect not only more recent information about volumes, but also new research findings will present further challenges. The best of our knowledge today suggests that in some instances volume may be a proxy for outcomes that patients experience. But it is only a proxy, and sometimes a poor one.

Few state governments have continued the hospital certificate of need (CON) programs popular in the 1970s. Even when such regulatory programs were popular, only a small proportion focused on the development of requirements for specific services or procedures. New York State is one of the very few remaining states that has an active CON program that does include requirements for specific services (e.g., CABG surgery, PTCA, organ transplantation). New

York's successful regionalization of CABG surgery services has been cited as responsible for severely limiting the number of low-volume hospitals performing this operation in the state and, together with its quality measurement and improvement program, for reductions in statewide mortality following the procedure (Grumbach 1995, Chassin 1996, Chassin 1997, Peterson 1998, Sollano 1999). There may be a role for such programs in other states and for other procedures. Political opposition to such regulation is widespread, however, and its implementation and operation require substantial organizational expertise and infrastructure.

Several recent initiatives have aimed to use private sector purchasing power to encourage managed care organizations to contract only with high volume-hospitals or to provide incentives to patients to seek care at such institutions. No published studies have evaluated the impact of such attempts to move substantial numbers of patients from low to high-volume hospitals. How strong such incentives would have to be for significant changes to occur is unknown. Previous research on the impact of publicly available data on mortality following CABG surgery documents the absence of significant movement of patients toward low-mortality hospitals (Chassin 1996), suggesting that information on outcome (or volume) alone may be insufficient to induce substantial changes in where patients receive care. A recent review documents the lack of use of data on performance of various parts of the health care system by consumers in their decision making (Marshall 2000).

Perhaps the most promising approach to the use of these data for improving outcomes is to employ them in quality improvement programs. To the extent that low volume (and poor performance) can be associated with underuse of effective processes of care, demonstrated to improve outcomes (or with an increased use of harmful practices), such knowledge can be put to direct use by these institutions to improve their performance. As a new generation of research produces more such data, the prospect of substantial improvement will increase. Among the most important research efforts to encourage are those which build prospectively-compiled, clinical databases that include all patients and all providers. Such databases can produce powerful risk-adjustment models that permit the production of valid outcome comparisons among providers and provide unparalleled opportunities to examine the effects of specific processes of care on carefully risk-adjusted outcomes. Government and private purchasers can accelerate this improvement by investing in the necessary research, by establishing and evaluating collaborative demonstrations, and by disseminating successful models.

ADDENDUM A: RATING THE QUALITY OF RESEARCH ON VOLUME AND OUTCOME

Objective of Scoring System: designed to measure the degree to which the study design is likely to reveal generalizable conclusions about the magnitude and nature of the relationship between volume and outcome.

Characteristic	Values	Scores
1. Representativeness of sample:	Not Representative	0 1
2. Number of hospitals or doctors	H < 20 and MD < 50 / H ≥120 or MD ≥50	0 1
3. Total sample size (cases):	<1000 / ≥1000	0 1
4. No. of Adverse Events	≤20 / 21-100 / >100	0 1 2
5. Unit of analysis:	Hospital or MD / Both separately / Both together / Both +	0 1 2 3
6. Appropriateness of patient selection	not measured / measured separately / measured and analyzed	0 1 2
7. Volume	2 categories / Multiple	0 1
8. Risk adjustment:	none / admin only / clinical data / clinical + C ≥.75 and H/L test +	0 1 2 3
9. Clinical processes of care:	not measured / One / 2+	0 1 2
10. Outcomes:	death only / death +	0 1

TOTAL POSSIBLE POINTS = 18

ADDENDUM B: FINAL MEDLINE SEARCH ALGORITHMS

(Trauma [Text word] OR transplant[mesh] OR transplantation[mesh] OR AIDS[mesh] OR liver diseases/surgery[mesh] OR lung diseases/surgery[MESH] OR coronary artery bypass[mesh] OR endarterectomy,carotid[mesh] OR neoplasms/surgery[mesh]OR angioplasty[mesh] OR myocardial infarction[mesh] OR vascular surgical procedures[mesh]OR peripheral vascular diseases/surgery[mesh]OR arthroplasty,replacement,knee[mesh]OR arthroplasty,replacement,hip[mesh]) AND utilization[subheading] AND (volume[Text word] OR frequency[Text word] OR frequent [Text word] OR statistics[subheading]) AND (outcome assessment[mesh] OR outcome and process assessment[mesh] OR outcome[Text word])

(Trauma [Text word] OR transplant[mesh] OR transplantation[mesh] OR AIDS[mesh] OR liver diseases/surgery[mesh] OR lung diseases/surgery[MESH] OR coronary artery bypass[mesh] OR endarterectomy,carotid[mesh] OR neoplasms [mesh]OR angioplasty[mesh] OR myocardial infarction[mesh] OR vascular surgical procedures[mesh]OR peripheral vascular diseases/surgery[mesh]OR arthroplasty,replacement,knee[mesh]OR arthroplasty,replacement,hip[mesh]) AND regionalization

ADDENDUM C: SUMMARY OF STUDIES OF VOLUME AND OUTCOME IN CABG SURGERY

CABG

Study	Population	Time Period	Patient #	MD #	Hosp #	Unit of Analysis	Primary Outcome	Risk Adjustment Data Source	Definition of Low Volume (per yr)	Volume:Outcomes Results	Score
Riley 1985	Medicare US	1979-1980	6,157	NS	990	Hosp	60 day death	Admin	Continuous	Model estimates of 60-day deaths: Vol Deaths 20 6.7% 80 5.6%	7
Showstack 1987	All CA	1983	18,996	NS	77	Hosp	Inpt death	Admin	Low:<101 hi:>350	RAMR:5.2 vs 3.1%	8
Hannan 1989	AllNY	1986	9,774	353	27	MD Hosp Both	Inpt death	Admin	MD:low<116 hi:>116 Hosp:low:<224 hi:>650	RAMRs: LVP HVP LVH 5.7% 5.5% HVH 3.8% 3.2% Total 4.9% 4.0%	10
Hannan 1991	All NY	1989	12,448	126	30	MD Hosp Both	Inpt mort	Clinical	MD:low:<55 hi:>259 Hosp:low:<200 hi:>849	RAMRs: MD: 8.14 vs 2.43% Hosp: 7.25 vs. 2.85% MD + Hosp: See Table 2.	11
Farley 1992	All US HCUP	1980-1989	146,890	NS	62	Hosp	Inpt death	Admin	Continuous	mortality w/ volume	8
Grumbach 1995	All NY,CA,Canada (Ont,Man,BC)	1987-1989	116,593	NS	157	Hosp	Inpt death w/in 14 days	Admin	Low:<100 hi:>500 Canada had no lowest vol hosps, few in 100-199.	RAMRs: NY: 4.1 vs 2.6% CA: 4.7 vs 2.4% CAN: No relationship.	8
Hannan 1995	All NY	1989-1992	57,187	126-136	30	MD	Inpt mort	Clin	low:<51 hi:>150	RAMRs over 4 yrs: LVP: 60% HVP: 34% LVS who stay low vol over 4 yrs had very high RAMRs but 12% left practice of CABG	9
Shroyer 1996	All cases at VA hosps doing CABG	1987-1992	23,986	NS	44	Hosp	30 day death	Clin C = .72	<100 No true HVH (only 3 bet 300-400)	No relationship	7
Sollano 1999	All NY	1990-1995	97,137	1205	31	Hosp	Inpt death	Admin	<100 Only 1 LVH	No relationship	8

Abbreviations:
RAMR: risk-adjusted mortality rate
LVP: low-volume physician
LVH: low-volume hospital
HVP: high-volume physician
HVH: high-volume hospital
NS: not specified

ADDENDUM D: SUMMARY OF STUDIES OF VOLUME AND OUTCOME IN PEDIATRIC CARDIAC SURGERY

PEDIATRIC CARDIAC SURGERY

Study	Population	Time Period	Patient #	MD #	Hosp #	Unit of Analysis	Primary Outcome	Risk Adjustment Data Source	Definition of Low Volume	Volume:Outcomes Results	Score
Jenkins 1995	All children CA, MA	1988-1989	2,833	NS	37	Hosp	Inpt death	Admin	Low:<10/yr hi:>300/yr	18.5 vs 3.0% HVH had higher complexity cases and mort of those cases	8
Hannan 1998a	All children NY	1992-1995	7,169	NS	16	MD Hospital Both	Inpt death	Clinical C=.818	MD:<75 Hosp:<100	RAMRs: LVP HVP Total LVH 8.94 7.45 8.26% HVH 8.47 5.45 5.95% Total 8.77 5.90	9
Sollano 1999	All children NY	1990-1995	7,199	NS	16	Hosp	Inpt death	Admin C=.617	Continuous	Adjusted OR=0.944 for vol/100 cases; Adjusted OR=.636 for neonates	6

Abbreviations:

RAMR: risk-adjusted mortality rate
LVP: low-volume physician
LVH: low-volume hospital
HVP: high-volume physician
HVH: high-volume hospital
NS: not specified

ADDENDUM E: SUMMARY OF STUDIES OF VOLUME AND OUTCOME IN CAROTID ENDARTERECTOMY

CAROTID ENDARTERECTOMY

Study	Population	Time Period	Patient #	MD #	Hospital #	Unit of Analysis	Primary Outcome	Risk Adjustment Data Source	Definition of Low Volume	Volume:Outcomes Results	Score
Brott, 1984	All Cincinnati	1980	431	47	16	MD	Inpt death, Stroke	None	MD: 5/yr	MD: No relationship	3
Kempczinski,1986	All Cincinnati	1983-1984	750	61	16	MD Hospital	Inpt death, stroke	None	MD: < 12/yr Hosp: < 50/yr	MD: No relationship Hosp: No relationship	6
Kirshner, 1989	All Rochester,NY	1984-1985	1035	22	6	MD Hospital	30 day death Stroke, MI	None	MD: 5/yr Hosp: NS	MD: No relationship Hosp: No relationship	6
Richardson, 1989	Medicare Kentucky	1983-1984	738	98	41	MD Hospital	Inpt death	None	MD: < 3/yr	MD: < 3/yr v. > 12/yr higher mortality (6.1% v. 2.3%) Hosp: No relationship	5
Fisher, 1989	Medicare New England	1984-1985	2089	NS	139	Hospital	30 day death	Admin	Hosp: 5/yr	Hosp: mortality/stroke (3.2% v. 1.1%)	5
Brook, 1990	Medicare 13 US areas	1981	1171	NS	NS	MD	30 day death, stroke, MI	Clinical data, appropriateness	MD: 3/yr	MD: 3/yr v. >37/yr More complications RR=0.9 (11.9% v. 10.7%)	11
Edwards, 1991	All Tennessee	1979-1986	11,199	190	NS	MD Hospital	Inpt death	None	MD: 1-12/yr Hosp: NS	MD: volume, mortality Hosp: No relationship	8
Segal, 1993	Medicare Penn	1989-1992	5657	652	163	MD Hospital	Inpt death	None	MD: < 14/yr Hosp:<14/yr	MD:< 14/yr v. 14/yr mortality (2.6%v.1.2%) Hosp: <14/yrv.≥14/yr Trend [0.8] (2.1%v.1.9%)	7
Ruby, 1996	All Connecticut	1985-1991	3997	226	32	MD	Inpt death, Stroke (abst)	None	MD: 1/yr	MD: 1/yr v. > 10/yr 2.5X mortality/ cva	7
Perler, 1998	All Maryland	1990-1995	9918	NS	48	Hospital	Inpt death	None	Hosp: 10/yr	Hosp: trend towards (.08) volume, mortality	5
Karp, 1998	Medicare Georgia	1993	1945	NS	67	Hospital	30 day death Stroke, MI	Admin appropriateness	Hosp: 10/yr	Hosp: 10/yr v. 50/yr 2.6X mortality/ cva	6
Hsia, 1998	Medicare US sample	1985-1996	63,137	NS	2607	Hospital	30 day death	None	Hosp:1-20/yr	Hosp: 1-20/yr v. 50/yr mortality (2.5%v.1.9%)	5
Manheim, 1998	All California	1982-1994	106,493	NS	NS	Hospital	Inpt death	Admin	Hosp: <20/yr	Hosp: <20/yr vs. > 100/yr mortality OR=.66 (1.2% v. 1.7%)	6
Wennberg, 1998	Medicare US sample	1992-1993	113,300	NS	2698	Hospital	30 day death	Admin	Hosp: 1-6/yr	Hosp: volume, mortality	6
Kantonen, 1998	All Finland	1991-1994	1600	104	23	MD Hospital	30 day death Stroke	Clinical	MD: < 10/yr	MD: volume, mortality Hosp: No relationship	10
Cebul, 1998	Medicare Ohio	1993-1994	678	115	478	MD Hospital Together	30 day death Stroke	Clinical Indication	MD:< 21/yr Hosp: 62/yr	MD:No relationship Hosp: 62/yr 29% death/ stroke	9
Hannan, 1998b	All New York	1990-1995	28,207	518	161	MD Hospital Together	Inpt death	Admin C=.715	MD: 1-4/yr Hosp: 100/yr	MD: 1-4/yr mortality Hosp: 100/yr mortality MD vol more important	10
Khuri, 1999	Veterans US sample	1991-1997	10,173	NS	93	Hospital	30 day death Stroke	Clinical C=.72	Hosp: <10/yr	Hosp: No relationship death or stroke	8

ADDENDUM F: SUMMARY OF STUDIES OF VOLUME AND OUTCOME IN ABDOMINAL AORTIC ANEURYSM SURGERY

ABDOMINAL AORTIC ANEURYSM

Study	Population	Time period	Patient #	MD #	Hospital #	Unit of analysis	Primary outcome	Risk adjustment Data source	Definition of low volume	Volume:Outcomes Results	Score
Hannan 1989	All NY	1986	1635	508	170	MD Hosp Both	Inpt death	Admin	MD Low:<=4 Hi: >4 Hosp Low:<=5 Hi:>35	Mortality related to surgeon and hosp vol; Surgeon vol more import. RAMRs: LVP HVP LVH 21.9% 18.7% HVH 19.7% 10.8% Total 20.3% 14.7%	10
Amundsen 1990	All AAA Trial Norway	1981-1983	444 total 279 elect.	NS	26	Hosp	Inpt death	None	1-9; Also measured volume of all vascular operations	Unruptured mortality related to unruptured vol (OR=2.7) and to all vascular vol (OR=2.6), but not to ruptured vol; Ruptured mortality related to all vascular vol (OR=2.5) but not to unruptured vol;	5
Hannan 1992a	All NY	1985-1987	1639 R 6042 U	363 to 477	155 to 169	MD Hosp Both	Inpt death	Admin Disease staging	MD:0-2 Hosp:0-9; Also measured vol of any aorta operation	Unruptured mortality related to hosp vol, not surgeon vol; Ruptured mortality related to surgeon vol, not hosp vol; Results same for vol of any aorta operation	10
Katz 1994	All MI	1980-1990	8185 U 1829 R	NS	176	Hosp	Inpt death	Admin	Low:<21 Hi: >=21 unruptured Low:<5 Hi:>=5 ruptured cases	Unruptured crude mortality: 8.9 vs 6.2%, (OR=1.2) Ruptured crude mortality: 53.6 vs 45.7%	7
Kazmers 1996	VA US	1991-1993	3687	NS	116	Hosp	Inpt death	Admin	<31	OR=0.6 for all AAA	6
Wen 1996	All Ontario	1988-1992	1203 R 5492 U	NS	78 rupt; 70 unrup	Hosp	Inpt death	Admin	<10 ruptured cases for 5 yrs;<50 unruptured cases for 5 yrs;	Unruptured AAA: OR=0.94; Ruptured AAA: no relationship	8

Rutledge 1996	All NC	1988-1993	1480 R 14138 total	_100	157	MD Hosp	Inpt death - ruptured AAA only	Admin	Continuous; Measured vol of ruptured cases and vol of unruptured cases	MD: Ruptured AAA mortality related to surgeon vol of ruptured cases but not to surgeon vol of unruptured cases; Hosp: Trend, but not significant relationship to hosp vol of ruptured or unruptured cases	8
Dardik 1998	All Maryland	1990-1995	527	226	45	MD Hosp	Inpt death - ruptured AAA only	Admin	MD: <10 ruptured cases Hosp: <10 ruptured; <50 unruptured	MD: Ruptured AAA mortality related to surgeon vol of ruptured cases; Hosp: Ruptured AAA mortality unrelated to hosp vol of ruptured or unruptured cases	8
Manheim 1998	All CA	1998	7327 ruptured 35130 unruptured	NS	58	Hosp	Inpt death	Admin	Low: <50 high: >=50	Unruptured: 78 vs 84% Ruptured: 74 vs 49%	8
Pronovost 1999	All Maryland	1994-1996	2987	NS	46	MD Hosp	Inpt death	Admin	MD:<8 Hosp:<36	MD:No relationship Hosp: OR=1.7	8
Sollano 1999	All NY	1990-1995	9847	NS	195	Hosp	Inpt death	Admin	<=14.7	OR=.782/100 cases;	8
Khuri 1999	VA US	1991-1997	3747	NS	123	Hosp	30 day death	Clinical C=.75	0-3 Also measured vol of all peripheral vascular procedures	No relationship	9

Abbreviations:

RAMR: risk-adjusted mortality rate
LVP: low-volume physician
LVH: low-volume hospital
HVP: high-volume physician
HVH: high-volume hospital
OR: odds ratio
NS: not specified

ADDENDUM G: SUMMARY OF STUDIES OF VOLUME AND OUTCOME IN CANCER SURGERY

BREAST

Study	Population	Time period	Patient #	MD #	Hospital #	Unit of analysis	Primary outcome	Risk adjustment Data source	Definition of low volume	Volume:Outcomes Results	Score
Roohan 1998	All women NY	1984-1989	47890	NS	266	Hosp	5 yr survival	Clinical	Hosp: Low: <10/yr high:>149/yr	OR=1.6	10
Sainsbury 1995	All women Yorkshire, UK	1979-1988	12861	180	NS	MD	5 yr survival	Clinical	MD: <30/yr	Vol / Adjusted RR Ratio <10 1.0 10-29 0.97 30-49 0.85 >=50 0.86	11

Abbreviations:
OR: odds ratio
RR: relative risk
NS: not specified

COLORECTAL

Study	Population	Time period	Patient #	MD #	Hospital #	Unit of analysis	Primary outcome	Risk adjustment Data source	Definition of low volume	Volume:Outcomes Results	Score
Hannan 2000	All NY	1994-1997	22128	2052	229	MD Hosp Both	Inpt death	Admin	MD:low:<12 high: >34 Hosp:low:<84 high:>253	RAMR for LVH 1.93% > HVH; No MD effect when hosp volume controlled	10
Harmon 1999	All Maryland	1992-1996	9739	812	50	MD Hosp Both	Inpt death	Admin	MD:<6/yr Hosp:<40/yr	MD:HVS v LVS;OR=.64 Hosp:HVH v LVH;OR=.78 MVS at HVH/MVP equiv to HVS; HVS better at any hosp	10
Parry 1999	All NW UK	1993 (6 mos)	927	123	39	MD Hosp	30- day death 3-year survival	Clinical	MD:<7 in 6 mos Hosp:<30 in 6 mos	No relationship	9
Gordon 1999	All Maryland Total colectomy	1989-1997	1015	NS	51	Hosp	Inpt death	Admin	Hosp: <10/yr	No relationship	8
Porter 1998	All Edmonton Rectal CA	1983-1990	683	52	5	MD	Local recurrence Disease specific survival	Clinical	MD: <21/yr	Local recurrence HR=1.8 DSS: HR=1.4 HVP no more likely to give adjuvant Rx; HVP more likely to do LAR	7

RAMR: risk-adjusted mortality rate
LVP: low-volume physician
LVH: low-volume hospital
HVP: high volume physician
HVH: high-volume hospital
MRP: medium-volume physician
DSS: disease specific survival
NS: not specified
HR: hazards ratio
LAR: low anterior resection

ESOPHAGUS

Study	Population	Time period	Patient #	MD #	Hospital #	Unit of analysis	Primary outcome	Risk adjustment Data source	Definition of low volume	Volume:Outcomes Results	Score
Gordon 1999	All Maryland (Benign and malignant)	1989-1997	518	NS	51	Hosp	Inpt death	Admin	Hosp: 10/yr Volume of 6 complex GI procedures	Vol RR 10 3.8 11-20 4.0 21-50 2.4 >200 1.0	6
Begg 1998	Medicare US	1984-1993	503	NS	190	Hosp	Inpt death	Clinical	Hosp: Low: 5/yr high: 11/yr	Mortality 17.3 vs 3.4%	6
Patti 1998	All CA	1990-1994	1561	NS	273	Hosp	Inpt death	Admin	Hosp: Low: 5/yr high:>30/yr	Mortality 17 vs 6%	8

RR: Relative Risk

LUNG

Study	Population	Time period	Patient #	MD #	Hospital #	Unit of analysis	Primary outcome	Risk adjustment Data source	Definition of low volume	Volume:Outcomes Results	Score
Hannan 2000	All NY Lobectomies	1994-1997	6954	373	178	MD Hosp Both	Inpt death	Admin	MD:<23/yr Hosp:<38/yr	Hosp: RAMR for LVH 1.65% > HVH MD: no relationship	10
Begg 1998	Medicare US Pneumo-nectomies	1984-1993	1375	NS	313	Hosp	30-day mortality	Clinical	Hosp: <6/yr	No relationship	8
Romano 1992	All CA All resections	1983-1986	12439	NS	389	Hosp	Inpt death	Admin	Hosp: <9/yr	Lesser resections (high relative to low volume): OR=0.6 Pneumonectomy: OR=0.6	8

Abbreviations:

LVP: low-volume physician
LVH: low-volume hospital
HVP: high-volume physician
HVH: high-volume hospital
RAMR: risk-adjusted mortality rate
OR: odds ratio

PANCREAS

Study	Population	Time period	Patient #	MD #	Hospital #	Unit of analysis	Primary outcome	Risk adjustment Data source	Definition of low volume	Volume:Outcomes Results	Score
Gordon 1999	All Maryland; Benign & malignant	1989-1997	1092	NS	51	Hosp	Inpt death	Admin	Hosp: 10/yr	Vol RR 10 12.5 11-20 10.4 21-50 6.3 >200 1	7
Birkmeyer 1999a	Medicare US Benign & malignant	1992-1995	7229	NS	1772	Hosp	3 yr death	Admin	Hosp Very low:<1 Low: 1-2 Hi: 5	OR=0.69	7
Birkmeyer 1999b	Medicare US Benign & malignant	1992-1995	7229	NS	1772	Hosp	Inpt death 30-d death	Admin	Hosp Very low:<1 Low: 1-2 Hi: 5	Inpt death: 16% v 4.1% (very high 1.7%) 30-d death: 12.9 v 3.0%	7
Sosa 1998a	All Maryland	1990-1995	1236	373	48	MD Hosp Both	Inpt death	Admin	MD: Low:<5, Hi:>50 Hosp: Low:<5 Hi: 20	LVH vs HVH:RR=19.3 HVH better, regardless of MD volume	9
Begg 1998	Medicare US	1984-1993	742	NS	252	Hosp	Inpt death	Clinical	Low:<6 high:>10/yr	Mortality 12.9 vs 5.8%	6
Simunovic 1999	All Ontario	1988/89 or 1994/95	842	NS	68	Hosp	Inpt death 64-d death	Admin	<22	LVH: OR=5.1 MVH: OR=4.5	6
Glasgow 1996	All CA	1990-1994	1705	NS	298	Hosp	Inpt death Bleeding Infection	Admin	Low:1-5 high:>50	RAMR 14%vs. 3.5%	8
Imperato 1996	Medicare NY	1991-94	579	NS	117	Hosp	Inpt death	Admin	Low:1-5/yr high:>25/yr	Mortality: 14.3 vs 2.2% (RR 6.87)	5
Wade 1996	Dept of Defense US	1989-1994	130	NS	111	Hosp	Inpt death	None	<1	Mortality <1: 6% 1-2: 9% >2: 9% (no p value given)	3
Lieberman 1995	All NY	1984-1991	1972	748	184	MD Hosp Both	Inpt death	Admin	MD:<9 Hosp:<10	MD: 6% vs 13%; Hosp: 5% vs 18.9%; Both: Only hospital volume is important	10
Gordon 1995	All Maryland	1988-1993	501	NS	39	Hosp	Inpt death	Admin	Low:<1-5/yr high:>20/yr	Mortality 19% vs 2.2% (RR=8.7)	6

OR: odds ratio; NS: not specified; RR: relative risk; and LVH: low-volume hospital.

CANCER MISCELLANEOUS

Study	Population	Time period	Patient #	MD #	Hospital #	Unit of analysis	Primary outcome	Risk adjustment Data source	Definition of low volume	Volume:Outcomes Results	Score
Hannan 2000	All NY Gastrectomy for Cancer	1994-1997	3711	1114	207	MD Hosp Both	Inpt death	Admin	MD:1-2 Hosp:1-15	Risk-adjusted increase in rate for lowest relative to highest volume quartile Hosp: 7.1% Surgeon: 5.7% No MD effect when hosp volume controlled.	10
Glasgow 1999	All CA **Hepatic resections for Cancer**	1990-1994	507	NS	138	Hosp	Inpt death	Admin	Low:<=2 high:>16	Risk-adjusted mortality rate: Low: 22.7 High: 9.4%	6
Gordon 1999	All Maryland Biliary tract anastomosis, gastrectomy, hepatic lobectomy (benign and malignant)	1989-1997	938; 705; 293	NS	51	Hosp	Inpt death	Admin	<11 Measured vol of 6 complex GI procedures	Biliary tract anastomosis:adjusted RR=5.3 Gastrectomy:no relationship Hepatic lobectomy:adjusted RR=4.7 6 GI procedures: Benign: no relationship Malignant:adjusted RR=5.2	6
Begg 1998	Medicare/US Pelvic exenteration, hepatic resection	1984-1993	1592; 801	NS	250+	Hosp	30-day death	Clinical	Low:<1-5 high:≥11	Unadjusted 30-day mortality: Pelvic:3.7 vs. 1.5% Hepatic:5.4 vs. 1.7%	7

Abbreviations:
LVP: low-volume physician
LVH: low-volume hospital
HVP: high-volume physician
HVH: high-volume hospital
NS: not specified

ADDENDUM H: SUMMARY OF STUDIES OF VOLUME AND OUTCOME IN CORONARY ANGIOPLASTY

CORONARY ANGIOPLASTY

Study	Population	Time period	Patient #	MD #	Hosp #	Unit of analysis	Primary outcome	Risk adjustment Data source	Definition of low volume	Volume:Outcomes Results	Score
Ritchie 1999	All US	1993-1994	163,527	NS	213	Hosp	Inpt death, CABG	Admin	Hosp: _200/yr	Hosp: _ vol _ death/CABG No assoc with death alone	9
Maynard 1999	All CA	1993, 1996	43,040	NS	120	Hosp	Inpat death, CABG	Admin	Hosp: <85/yr	Hosp: _ vol _ CABG No assoc with death alone	9
Malenka 1999	All ME,NH,VT	1994-1996	15080	47	6	MD	Inpt death, MI CABG, Angio/clin. success	Clinical	MD: <84/yr	MD: trend_ vol _ death(.09) trend _ vol _ MI (.065) No relationship with CABG	9
McGrath 1998	All ME,NH,VT	1990-1993	12488	31	5	MD	Inpt death, CABG, MI Angio/clinical success	Clinical	MD: _85/yr	MD: _ vol _ CABG, Trend _ vol _ MI (.06) No assoc with death alone _ vol _ angio/clin. success	7
Hannan 1997a	All NY	1991-1994	62,670	130	31	MD Hosp Both	Inpt death CABG	Clinical	MD:<75/yr Hosp:<400/yr	MD: <75/yr _ death,_CABG Hosp:<400/yr_death,_CABG	13
Jollis 1997	Medicare US	1992	119,886	6115	984	MD Hosp Both	Inpt death 30-day death Inpt CABG	Admin	MD: <25/yr Hosp:<100/yr	MD: _ vol _ death/CABG No assoc with death alone Hosp:_ vol _ death, _CABG, and _death/CABG	10
Philips 1995	All CA	1989	24,856	NS	110	Hosp	Inpt death CABG	Admin	Hosp: _101/yr	Hosp: _ vol_ death/CABG	9
Jollis 1994	Medicare US	1987-1990	217,836	NS	1194	Hosp	Inpt death 30-day death Inpt CABG	Admin	Hosp: _46/yr	Hosp: _ vol _ death _ vol _ CABG	8
Ritchie 1993	All CA	1989	24,883	NS	110	Hosp	Inpt death, CABG	Admin	Hosp: _200/yr	Hosp: _ vol _ death/CABG _ vol _CABG No assoc with death alone	9

ADDENDUM I: SUMMARY OF STUDIES OF VOLUME AND OUTCOME IN ACUTE MYOCARDIAL INFARCTION

AMI

Study	Population	Time period	Patient #	MD #	Hospital #	Unit of analysis	Primary outcome	Risk adjustment Data source	Definition of low volume	Volume:Outcomes Results	Score
Thiemann 1998	Medicare US	1994-1995	98898	NS	NS	Hosp	30 day &1 yr survival	Clinical Process of care Specialty Model C=.79	Hosp: <1.4/wk	Hosp:_ volume _ mortality _ volume by 5.5 Pts/wk, _ 30-survival HR=1.17 _ 1 yr survival HR=1.05 Measured processes of care account for 1/3 of survival benefit	11
Casale 1998	All PA	1993	30715	NS	NS	MD	Inpt death	Clinical Specialty Model C=.881	MD: _12/yr	MD: high volume (>12/yr) _ mortality OR=.89	9
Farley 1992	All US (HCUP)	1980-1987	974,803	NS	426	Hosp	Inpt death	Admin	Continuous	Hosp: 10%_ volume, 2.2%_ inpt mortality No selective referral effect	8

ADDENDUM J: SUMMARY OF STUDIES OF VOLUME AND OUTCOME IN AIDS

AIDS

Study	Population	Time period	Patient #	MD #	Hospital #	Unit of analysis	Primary outcome	Risk adjustment Data source	Definition of low volume	Volume:Outcomes Results	Score
Bennett 1989	All Metropolitan CA	1986-1987	257	NS	15	Hosp	Inpt death	Admin	<30 AIDS admissions per 10,000 discharges	LVH 33% HVH 12%	2
Stone 1992	All MA	1987-1988	3007	NS	40	Hosp	30 day death	Clinical	<43 AIDS admissions per 10,000 discharges	Adjusted RR 2.92 (low vs high)	7
Turner 1992	HCUP	1986-1987	10538	NS	258	Hosp	Inpt death	Admin	Continuous	4.3%_ in odds of death for each 100-case _ in hosp vol.	8
Bennett 1992	All NYC	1987	3126	NS	73	Hosp	Inpt death 30 day death	Admin	80 PCP admissions	Vol OR <80 1.0 0.64 >239 0.67	8
Kitahata 1996	Group Health WA	1984-1994	403	125	NS	MD	Death since dx	Clinical	low:<2 high: >5 cases per MD	OR =0.57	6
Hogg 1998	All Canada	1987-1994	38075	NS	513	Hosp	Inpt death	Admin	low:<1 high:>100 admissions per year	OR=0.64	8

LVH: low-volume hospital
HVH: high-volume hospital
RR: relative risk
OR: odds ratio
NS: not specified

Bibliography

1. **Aass N, Klepp O, Cavallin-Stahl E, et al.** Prognostic factors in unselected patients with nonseminomatous metastatic testicular cancer: a multicenter experience. *Journal of Clinical Oncology*. 1991;9(5):818-26.
2. **Adam DJ, Mohan IV, Stuart WP, Bain M, Bradbury AW.** Community and hospital outcome from ruptured abdominal aortic aneurysm within the catchment area of a regional vascular surgical service. *J Vasc Surg*. 1999;30(5):922-8.
3. **Adams DF, Fraser DB, Abrams HL.** The complications of coronary arteriography. *Circulation*. 1973;48(3):609-18.
4. **al-Haider AS, Wan TT.** Modeling organizational determinants of hospital mortality. *Health Serv Res*. 1991;26(3):303-23.
5. **Alberts KA, Brismar B, Nygren A.** Major differences in trauma care between hospitals in Sweden: a preliminary report. *Qual Assur Health Care*. 1993;5(1):13-7.
6. **Amundsen S, Skjaerven R, Trippestad A, Soreide O.** Abdominal aortic aneurysms. Is there an association between surgical volume, surgical experience, hospital type and operative mortality? Members of the Norwegian Abdominal Aortic Aneurysm Trial. *Acta Chirurgica Scandinavica*. 1990;156(4):323-7; discussion 327-8.
7. **Anderson GM, Lomas J.** Regionalization of coronary artery bypass surgery. Effects on access. *Med Care*. 1989;27(3):288-96.
8. **Andren-Sandberg A, Ihse I.** Factors influencing survival after total pancreatectomy in patients with pancreatic cancer. *Ann Surg*. 1983;198(5):605-10.
9. **Banta D, Bos M.** The relation between quantity and quality with coronary artery bypass graft (CABG) surgery. *Health Policy*. 1991;18(1):1-10.
10. **Banta HD, Engel GL, Schersten T.** Volume and outcome of organ transplantation. *Int J Technol Assess Health Care*. 1992;8(3):490-505.
11. **Barbash GI, White HD, Modan M, et al.** Outcome of thrombolytic therapy in relation to hospital size and invasive cardiac services. The Investigators of the International Tissue Plasminogen Activator/Streptokinase Mortality Trial. *Arch Intern Med*. 1994;154(19):2237-42.
12. **Becker RM.** Outcome versus volume in coronary bypass operations. *Ann Thorac Surg*. 1997;63(3):917-8.
13. **Begg CB, Cramer LD, Hoskins WJ, Brennan MF.** Impact of hospital volume on operative mortality for major cancer surgery. *JAMA*. 1998;280(20):1747-51.
14. **Bennett CL, Garfinkle JB, Greenfield S, et al.** The relation between hospital experience and in-hospital mortality for patients with AIDS-related PCP. *JAMA*. 1989;261(20):2975-9.

15. **Bennett CL, Gertler P, Guze PA, Garfinkle JB, Kanouse DE, Greenfield S.** The relation between resource use and in-hospital mortality for patients with acquired immunodeficiency syndrome-related Pneumocystis carinii pneumonia. *Arch Intern Med*. 1990;150(7):1447-52.
16. **Bennett CL, Adams J, Gertler P, et al.** Relation between hospital experience and in-hospital mortality for patients with AIDS-related Pneumocystis carinii pneumonia: experience from 3,126 cases in New York City in 1987. *J Acquir Immune Defic Syndr*. 1992;5(9):856-64.
17. **Birkmeyer JD, Sharp SM, Finlayson SR, Fisher ES, Wennberg JE.** Variation profiles of common surgical procedures. *Surgery*. 1998;124(5):917-23.
18. **Birkmeyer JD, Warshaw AL, Finlayson SR, Grove MR, Tosteson AN.** Relationship between hospital volume and late survival after pancreaticoduodenectomy. *Surgery*. 1999a;126(2):178-83.
19. **Birkmeyer JD, Finlayson SR, Tosteson AN, Sharp SM, Warshaw AL, Fisher ES.** Effect of hospital volume on in-hospital mortality with pancreaticoduodenectomy. *Surgery*. 1999b;125(3):250-6.
20. **Black N, Johnston A.** Volume and outcome in hospital care: evidence, explanations and implications. *Health Serv Manage Res*. 1990;3(2):108-14.
21. **Brook RH, Park RE, Chassin MR, Kosecoff J, Keesey J, Solomon DH.** Carotid endarterectomy for elderly patients: predicting complications. *Ann Intern Med*. 1990a;113(10):747-53.
22. **Brook RH, Park RE, Chassin MR, Solomon DH, Keesey J, Kosecoff J.** Predicting the appropriate use of carotid endarterectomy, upper gastrointestinal endoscopy, and coronary angiography. *N Engl J Med*. 1990b;323(17):1173-7.
23. **Brott T, Thalinger K.** The practice of carotid endarterectomy in a large metropolitan area. *Stroke*. 1984;15(6):950-5.
24. **Bunker JP, Luft HS, Enthoven A.** Should surgery be regionalized? *Surg Clin North Am*. 1982;62(4):657-68.
25. **Casale PN, Jones JL, Wolf FE, Pei Y, Eby LM.** Patients treated by cardiologists have a lower in-hospital mortality for acute myocardial infarction. *Journal of the American College of Cardiology*. 1998;32(4):885-9.
26. **Cebul RD.** Quality improvement, case mix, and preventive procedures: when should the playing field be leveled for carotid endarterectomy? *QRB Qual Rev Bull*. 1993;19(5):150-1.
27. **Cebul RD, Snow RJ, Pine R, Hertzer NR, Norris DG.** Indications, outcomes, and provider volumes for carotid endarterectomy. *JAMA*. 1998;279(16):1282-7.
28. **Chassin MR, Park RE, Lohr KN, Keesey J, Brook RH.** Differences among hospitals in Medicare patient mortality. *Health Serv Res*. 1989;24(1):1-31.
29. **Chassin MR, Hannan EL, DeBuono BA.** Benefits and hazards of reporting medical outcomes publicly. *N Engl J Med*. 1996;334(6):394-8.
30. **Chassin MR.** Assessing strategies for quality improvement. *Health Aff (Millwood)*. 1997;16(3):151-61.
31. **Chassin MR.** Appropriate use of carotid endarterectomy. *N Engl J Med*. 1998;339(20):1468-71.
32. **Chen H, Zeiger MA, Gordon TA, Udelsman R.** Parathyroidectomy in Maryland: effects of an endocrine center. *Surgery*. 1996;120(6):948-52; discussion 952-3.
33. **Cho YW, Cecka JM.** Center effect in the UNOS Renal Transplant Registry. *Clin Transpl*. 1992:333-46.
34. **Choti MA, Bowman HM, Pitt HA, et al.** Should hepatic resections be performed at high-volume referral centers? *J Gastrointest Surg*. 1998;2(1):11-20.
35. **Clark RE.** Outcome as a function of annual coronary artery bypass graft volume. The Ad Hoc Committee on Cardiac Surgery Credentialing of The Society of Thoracic Surgeons. *Ann Thorac Surg*. 1996;61(1):21-6.
36. **Clemmer TP, Orme JF, Jr., Thomas FO, Brooks KA.** Outcome of critically injured patients treated at Level I trauma centers versus full-service community hospitals. *Critical Care Medicine*. 1985;13(10):861-3.
37. **Codman E.** The product of a hospital. *Surgery Obstetrics and Gynecology*. 1914(April):491-6.

38. **Collette L, Sylvester RJ, Stenning SP, et al.** Impact of the treating institution on survival of patients with "poor-prognosis" metastatic nonseminoma. European Organization for Research and Treatment of Cancer Genito-Urinary Tract Cancer Collaborative Group and the Medical Research Council Testicular Cancer Working Party. *J Natl Cancer Inst*. 1999;91(10):839-46.
39. **Cooper A, Barlow B, DiScala C, String D, Ray K, Mottley L.** Efficacy of pediatric trauma care: results of a population-based study. *J Pediatr Surg*. 1993;28(3):299-303; discussion 304-5.
40. **Cooper A, Hannan EL, Bessey PQ, Farrell LS, Cayten CG, Mottley L.** An examination of the volume-mortality relationship for New York State trauma centers. *J Trauma*. 2000;48(1):16-23; discussion 23-4.
41. **Cordero L, Backes CR, Zuspan FP.** Very low-birth weight infant. I. Influence of place of birth on survival. *American Journal of Obstetrics and Gynecology*. 1982;143(5):533-7.
42. **Coyte PC, Hawker G, Croxford R, Wright JG.** Rates of revision knee replacement in Ontario, Canada. *J Bone Joint Surg Am*. 1999;81(6):773-82.
43. **Culler SD, Holmes AM, Gutierrez B.** Expected hospital costs of knee replacement for rural residents by location of service. *Medical Care*. 1995;33(12):1188-209.
44. **Dardik A, Burleyson GP, Bowman H, et al.** Surgical repair of ruptured abdominal aortic aneurysms in the state of Maryland: factors influencing outcome among 527 recent cases. *Journal of Vascular Surgery*. 1998;28(3):413-20; discussion 420-1.
45. **de Belder MA.** Cardiac intervention procedures in the United Kingdom 1997: developments in data collection. Council of the British Cardiovascular Intervention Society. *Heart*. 1999;82 Suppl 2:II2-9.
46. **Dragalin D, Perkins D, Plocher DW.** Institutes of Quality: Prudential's approach to outcomes management for specialty procedures. *QRB Qual Rev Bull*. 1990;16(3):111-5.
47. **Duffner PK, Cohen ME, Flannery JT.** Referral patterns of childhood brain tumors in the state of Connecticut. *Cancer*. 1982;50(8):1636-40.
48. **Edge SB, Schmieg RE, Jr., Rosenlof LK, Wilhelm MC.** Pancreas cancer resection outcome in American University centers in 1989-1990. *Cancer*. 1993;71(11):3502-8.
49. **Edwards WH, Morris JA, Jr., Jenkins JM, Bass SM, MacKenzie EJ.** Evaluating quality, cost-effective health care. Vascular database predicated on hospital discharge abstracts. *Annals of Surgery*. 1991;213(5):433-8.
50. **Edwards EB, Roberts JP, McBride MA, Schulak JA, Hunsicker LG.** The effect of the volume of procedures at transplantation centers on mortality after liver transplantation. *N Engl J Med*. 1999;341(27):2049-53.
51. **Ellerbeck EF, Jencks SF, Radford MJ, et al.** Quality of care for Medicare patients with acute myocardial infarction. A four-state pilot study from the Cooperative Cardiovascular Project. *JAMA*. 1995;273(19):1509-14.
52. **Ellis SG, Weintraub W, Holmes D, Shaw R, Block PC, King SB, 3rd.** Relation of operator volume and experience to procedural outcome of percutaneous coronary revascularization at hospitals with high interventional volumes. *Circulation*. 1997;95(11):2479-84.
53. **Espehaug B, Havelin LI, Engesaeter LB, Vollset SE.** The effect of hospital-type and operating volume on the survival of hip replacements. A review of 39,505 primary total hip replacements reported to the Norwegian Arthroplasty Register, 1988-1996. *Acta Orthop Scand*. 1999;70(1):12-8.
54. **Evans SM, Adam DJ, Murie JA, Jenkins AM, Ruckley CV, Bradbury AW.** Training in abdominal aortic aneurysm (AAA) repair: 1987-1997. *Eur J Vasc Endovasc Surg*. 1999;18(5):430-3.
55. **Farley DE, Ozminkowski RJ.** Volume-outcome relationships and in-hospital mortality: the effect of changes in volume over time. *Med Care*. 1992;30(1):77-94.
56. **Feit F, Mueller HS, Braunwald E, et al.** Thrombolysis in Myocardial Infarction (TIMI) phase II trial: outcome comparison of a "conservative strategy" in community versus tertiary hospitals. The TIMI Research Group. *J Am Coll Cardiol*. 1990;16(7):1529-34.
57. **Feuer EJ, Sheinfeld J, Bosl GJ.** Does size matter? Association between number of patients treated and patient outcome in metastatic testicular cancer. *J Natl Cancer Inst*. 1999;91(10):816-8.

58. **Field D, Hodges S, Mason E, Burton P.** Survival and place of treatment after premature delivery. *Archives of Disease in Childhood*. 1991;66(4 Spec No):408-10; discussion 410-1.
59. **Fink A, Yano EM, Brook RH.** The condition of the literature on differences in hospital mortality. *Med Care*. 1989;27(4):315-36.
60. **Finlayson SR, Birkmeyer JD, Tosteson AN, Nease RF, Jr.** Patient preferences for location of care: implications for regionalization. *Medical Care*. 1999;37(2):204-9.
61. **Fisher ES, Malenka DJ, Solomon NA, Bubolz TA, Whaley FS, Wennberg JE.** Risk of carotid endarterectomy in the elderly. *American Journal of Public Health*. 1989;79(12):1617-20.
62. **Flood AB, Ewy W, Scott WR, Forrest WH, Jr., Brown BW, Jr.** The relationship between intensity and duration of medical services and outcomes for hospitalized patients. *Med Care*. 1979;17(11):1088-102.
63. **Flood AB, Scott WR, Ewy W.** Does practice make perfect? Part I: The relation between hospital volume and outcomes for selected diagnostic categories. *Medical Care*. 1984a;22(2):98-114.
64. **Flood AB, Scott WR, Ewy W.** Does practice make perfect? Part II: The relation between volume and outcomes and other hospital characteristics. *Med Care*. 1984b;22(2):115-25.
65. **Frances CD, Go AS, Dauterman KW, et al.** Outcome following acute myocardial infarction: are differences among physician specialties the result of quality of care or case mix? *Arch Intern Med*. 1999;159(13):1429-36.
66. **Freeland MS, Hunt SS, Luft HS.** Selective contracting for hospital care based on volume, quality, and price: prospects, problems, and unanswered questions. *J Health Polit Policy Law*. 1987;12(3):409-26.
67. **Garnick DW, Luft HS, McPhee SJ, Mark DH.** Surgeon volume vs hospital volume: which matters more? *JAMA*. 1989;262(4):547-8.
68. **Gilchrist IC, Gardner LH, Muhlestein JB, et al.** Effect of institutional volume and academic status on outcomes of coronary interventions: the IMPACT-II experience. *Am Heart J*. 1999;138(5 Pt 1):976-82.
69. **Gillis CR, Hole DJ.** Survival outcome of care by specialist surgeons in breast cancer: a study of 3786 patients in the west of Scotland. *BMJ*. 1996;312(7024):145-8.
70. **Glasgow RE, Mulvihill SJ.** Hospital volume influences outcome in patients undergoing pancreatic resection for cancer. *Western Journal of Medicine*. 1996;165(5):294-300.
71. **Glasgow RE, Showstack JA, Katz PP, Corvera CU, Warren RS, Mulvihill SJ.** The relationship between hospital volume and outcomes of hepatic resection for hepatocellular carcinoma. *Arch Surg*. 1999;134(1):30-5.
72. **Gordon TA, Burleyson GP, Tielsch JM, Cameron JL.** The effects of regionalization on cost and outcome for one general high-risk surgical procedure. *Annals of Surgery*. 1995;221(1):43-9.
73. **Gordon TA, Bowman HM, Tielsch JM, Bass EB, Burleyson GP, Cameron JL.** Statewide regionalization of pancreaticoduodenectomy and its effect on in-hospital mortality. *Annals of Surgery*. 1998;228(1):71-8.
74. **Gordon TA, Bowman HM, Bass EB, et al.** Complex gastrointestinal surgery: impact of provider experience on clinical and economic outcomes. *J Am Coll Surg*. 1999;189(1):46-56.
75. **Goss ME.** Comments on "The condition of the literature on differences in hospital mortality". *Med Care*. 1990;28(7):563-6.
76. **Grassman ED, Johnson SA, Krone RJ.** Predictors of success and major complications for primary percutaneous transluminal coronary angioplasty in acute myocardial infarction. An analysis of the 1990 to 1994 Society for Cardiac Angiography and Interventions registries. *Journal of the American College of Cardiology*. 1997;30(1):201-8.
77. **Gray HH.** Cardiac interventional procedures in the UK 1992 to 1996. Council of the British Cardiovascular Intervention Society. *Heart*. 1999;82 Suppl 2:II10-7.
78. **Grover FL, Johnson RR, Shroyer AL, Marshall G, Hammermeister KE.** The Veterans Affairs Continuous Improvement in Cardiac Surgery Study. *Ann Thorac Surg*. 1994;58(6):1845-51.

79. **Grumbach K, Anderson GM, Luft HS, Roos LL, Brook R.** Regionalization of cardiac surgery in the United States and Canada. Geographic access, choice, and outcomes. *JAMA*. 1995;274(16):1282-8.
80. **Hamilton SM, Letourneau S, Pekeles E, Voaklander D, Johnston DW.** The impact of regionalization on a surgery program in the Canadian health care system. *Arch Surg*. 1997;132(6):605-9; discussion 609-11.
81. **Hamilton BH, Hamilton VH.** Estimating surgical volume--outcome relationships applying survival models: accounting for frailty and hospital fixed effects. *Health Econ*. 1997;6(4):383-95.
82. **Hamilton BH, Ho V.** Does practice make perfect? Examining the relationship between hospital surgical volume and outcomes for hip fracture patients in Quebec. *Med Care*. 1998;36(6):892-903.
83. **Hannan EL, O'Donnell JF, Kilburn H, Jr., Bernard HR, Yazici A.** Investigation of the relationship between volume and mortality for surgical procedures performed in New York State hospitals. *JAMA*. 1989;262(4):503-10.
84. **Hannan EL, Kilburn H, Jr., Bernard H, O'Donnell JF, Lukacik G, Shields EP.** Coronary artery bypass surgery: the relationship between inhospital mortality rate and surgical volume after controlling for clinical risk factors. *Medical Care*. 1991;29(11):1094-107.
85. **Hannan EL, Kilburn H, Jr., O'Donnell JF, et al.** A longitudinal analysis of the relationship between in-hospital mortality in New York State and the volume of abdominal aortic aneurysm surgeries performed. *Health Serv Res*. 1992a;27(4):517-42.
86. **Hannan E.** Provider Volume-Patient Outcome Relationships in the Provision of Medical Care: An Update. 1992b (unpublished).
87. **Hannan EL, Kilburn H, Jr., Racz M, Shields E, Chassin MR.** Improving the outcomes of coronary artery bypass surgery in New York State. *JAMA*. 1994;271(10):761-6.
88. **Hannan EL, Siu AL, Kumar D, Kilburn H, Jr., Chassin MR.** The decline in coronary artery bypass graft surgery mortality in New York State. The role of surgeon volume. *JAMA*. 1995;273(3):209-13.
89. **Hannan EL, Racz M, Ryan TJ, et al.** Coronary angioplasty volume-outcome relationships for hospitals and cardiologists. *JAMA*. 1997a;277(11):892-8.
90. **Hannan EL, Racz MJ, Jollis JG, Peterson ED.** Using Medicare claims data to assess provider quality for CABG surgery: does it work well enough? *Health Serv Res*. 1997b;31(6):659-78.
91. **Hannan EL, Stone CC, Biddle TL, DeBuono BA.** Public release of cardiac surgery outcomes data in New York: what do New York state cardiologists think of it? *Am Heart J*. 1997c;134(6):1120-8.
92. **Hannan EL, Racz M, Kavey RE, Quaegebeur JM, Williams R.** Pediatric cardiac surgery: the effect of hospital and surgeon volume on in-hospital mortality. *Pediatrics*. 1998a;101(6):963-9.
93. **Hannan EL, Popp AJ, Tranmer B, Fuestel P, Waldman J, Shah D.** Relationship between provider volume and mortality for carotid endarterectomies in New York State. *Stroke*. 1998b;29:2292-7.
94. **Hannan EL.** The relation between volume and outcome in health care. *N Engl J Med*. 1999a;340(21):1677-9.
95. **Hannan EL.** Percutaneous coronary interventions: heed the American College of Cardiology volume recommendations, but strive to improve quality. *J Am Coll Cardiol*. 1999b;34(5):1481-3.
96. **Hannan EL, Imperato PJ, Nenner RP, Starr H.** Laparoscopic and open cholecystectomy in New York State: mortality, complications, and choice of procedure. *Surgery*. 1999c;125(2):223-31.
97. **Hannan E.** The influence of hospital and surgeon volume on mortality for common cancer procedures. 2000b (in press).
98. **Hansing CE.** The risk and cost of coronary angiography. II. The risk of coronary angiography in Washington State. *JAMA*. 1979;242(8):735-8.
99. **Harmon JW, Tang DG, Gordon TA, et al.** Hospital volume can serve as a surrogate for surgeon volume for achieving excellent outcomes in colorectal resection. *Ann Surg*. 1999;230(3):404-11; discussion 411-3.

100. **Hartz AJ, Kuhn EM, Krakauer H.** The relationship of the value of outcome comparisons to the number of patients per provider. *Int J Qual Health Care*. 1997;9(4):247-54.
101. **Hermanek P, Hohenberger W.** The importance of volume in colorectal cancer surgery. *Eur J Surg Oncol*. 1996;22(3):213-5.
102. **Hertzer NR, Avellone JC, Farrell CJ, et al.** The risk of vascular surgery in a metropolitan community. With observations on surgeon experience and hospital size. *J Vasc Surg*. 1984;1(1):13-21.
103. **Hewitt M, Simone J.** (Institute of Medicine and Commission on Life Sciences, National Research Council). Ensuring Quality Cancer Care. 1999.
104. **Higashihara E, Baba S, Nakagawa K, et al.** Learning curve and conversion to open surgery in cases of laparoscopic adrenalectomy and nephrectomy. *J Urol*. 1998;159(3):650-3.
105. **Hillner BE, Smith TJ.** Hospital volume and patient outcomes in major cancer surgery: a catalyst for quality assessment and concentration of cancer services. *JAMA*. 1998;280(20):1783-4.
106. **Hirsch N, Hailey D, Martin C.** The evolution of heart, lung and liver transplantation services in Australia. *Health Policy*. 1995;34(1):63-71.
107. **Hogg RS, Raboud J, Bigham M, Montaner JS, O'Shaughnessy M, Schechter MT.** Relation between hospital HIV/AIDS caseload and mortality among persons with HIV/AIDS in Canada. *Clinical and Investigative Medicine. Medecine Clinique et Experimentale*. 1998;21(1):27-32.
108. **Horbar JD, Badger GJ, Lewit EM, Rogowski J, Shiono PH.** Hospital and patient characteristics associated with variation in 28-day mortality rates for very low birth weight infants. Vermont Oxford Network. *Pediatrics*. 1997;99(2):149-56.
109. **Horowitz MM, Przepiorka D, Champlin RE, et al.** Should HLA-identical sibling bone marrow transplants for leukemia be restricted to large centers? *Blood*. 1992;79(10):2771-4.
110. **Hosenpud JD, Breen TJ, Edwards EB, Daily OP, Hunsicker LG.** The effect of transplant center volume on cardiac transplant outcome. A report of the United Network for Organ Sharing Scientific Registry. *JAMA*. 1994;271(23):1844-9.
111. **Houghton A.** Variation in outcome of surgical procedures. *Br J Surg*. 1994;81(5):653-60.
112. **Hsia DC, Moscoe LM, Krushat WM.** Epidemiology of carotid endarterectomy among Medicare beneficiaries: 1985-1996 update. *Stroke*. 1998;29(2):346-50.
113. **Hughes RG, Hunt SS, Luft HS.** Effects of surgeon volume and hospital volume on quality of care in hospitals. *Medical Care*. 1987;25(6):489-503.
114. **Hughes RG, Garnick DW, Luft HS, McPhee SJ, Hunt SS.** Hospital volume and patient outcomes. The case of hip fracture patients. *Medical Care*. 1988;26(11):1057-67.
115. **Hulka F, Mullins RJ, Mann NC, et al.** Influence of a statewide trauma system on pediatric hospitalization and outcome. *J Trauma*. 1997;42(3):514-9.
116. **Hunt J, Hill D, Besser M, West R, Roncal S.** Outcome of patients with neurotrauma: the effect of a regionalized trauma system. *Aust N Z J Surg*. 1995;65(2):83-6.
117. **Iezzoni LI**. Risk adjustment for measuring healthcare outcomes. *Health Administration Press*. Chicago, IL. 1997.
118. **Imperato PJ, Nenner RP, Starr HA, Will TO, Rosenberg CR, Dearie MB.** The effects of regionalization on clinical outcomes for a high risk surgical procedure: a study of the Whipple procedure in New York State. *American Journal of Medical Quality*. 1996;11(4):193-7.
119. **International Neonatal Network .** The CRIB (clinical risk index for babies) score: a tool for assessing initial neonatal risk and comparing performance of neonatal intensive care units. [published erratum appears in Lancet 1993 Sep 4;342(8871):626]. *Lancet*. 1993;342(8865):193-8.
120. **Jenkins KJ, Newburger JW, Lock JE, Davis RB, Coffman GA, Iezzoni LI.** In-hospital mortality for surgical repair of congenital heart defects: preliminary observations of variation by hospital caseload. *Pediatrics*. 1995;95(3):323-30.
121. **Jollis JG, Peterson ED, DeLong ER, et al.** The relation between the volume of coronary angioplasty procedures at hospitals treating Medicare beneficiaries and short-term mortality. *New England Journal of Medicine*. 1994;331(24):1625-9.

122. **Jollis JG, DeLong ER, Peterson ED, et al.** Outcome of acute myocardial infarction according to the specialty of the admitting physician. *N Engl J Med*. 1996;335(25):1880-7.
123. **Jollis JG, Peterson ED, Nelson CL, et al.** Relationship between physician and hospital coronary angioplasty volume and outcome in elderly patients. *Circulation*. 1997;95(11):2485-91.
124. **Jones J, Rowan K.** Is there a relationship between the volume of work carried out in intensive care and its outcome? *International Journal of Technology Assessment in Health Care*. 1995;11(4):762-9.
125. **Jubelirer RA, Agarwal NN, Beyer FCd, et al.** Pediatric trauma triage: review of 1,307 cases. *J Trauma*. 1990;30(12):1544-7.
126. **Kagan RJ, Baker RJ.** The impact of the volume of neurotrauma experience on mortality after head injury. *American Surgeon*. 1994;60(6):394-400.
127. **Kantonen I, Lepantalo M, Salenius JP, Matzke S, Luther M, Ylonen K.** Influence of surgical experience on the results of carotid surgery. The Finnvasc Study Group. *European Journal of Vascular and Endovascular Surgery*. 1998;15(2):155-60.
128. **Karp HR, Flanders WD, Shipp CC, Taylor B, Martin D.** Carotid endarterectomy among Medicare beneficiaries: a statewide evaluation of appropriateness and outcome. *Stroke*. 1998;29(1):46-52.
129. **Kastrati A, Neumann FJ, Schomig A.** Operator volume and outcome of patients undergoing coronary stent placement. *J Am Coll Cardiol*. 1998;32(4):970-6.
130. **Katz DJ, Stanley JC, Zelenock GB.** Operative mortality rates for intact and ruptured abdominal aortic aneurysms in Michigan: an eleven-year statewide experience. *J Vasc Surg*. 1994;19(5):804-15; discussion 816-7.
131. **Kazmers A, Jacobs L, Perkins A, Lindenauer SM, Bates E.** Abdominal aortic aneurysm repair in Veterans Affairs medical centers. *Journal of Vascular Surgery*. 1996;23(2):191-200.
132. **Keeler EB, Rubenstein LV, Kahn KL, et al.** Hospital characteristics and quality of care. *JAMA*. 1992;268(13):1709-14.
133. **Kelly JV, Hellinger FJ.** Physician and hospital factors associated with mortality of surgical patients. *Med Care*. 1986;24(9):785-800.
134. **Kelly JV, Hellinger FJ.** Heart disease and hospital deaths: an empirical study. *Health Services Research*. 1987;22(3):369-95.
135. **Kempczinski RF, Brott TG, Labutta RJ.** The influence of surgical specialty and caseload on the results of carotid endarterectomy. *J Vasc Surg*. 1986;3(6):911-6.
136. **Khuri SF, Daley J, Henderson W, et al.** Relation of surgical volume to outcome in eight common operations: results from the VA National Surgical Quality Improvement Program. *Ann Surg*. 1999;230(3):414-29; discussion 429-32.
137. **Kimmel SE, Berlin JA, Laskey WK.** The relationship between coronary angioplasty procedure volume and major complications. *JAMA*. 1995;274(14):1137-42.
138. **Kimmel SE, Kolansky DM.** Operator volume as a "risk factor". *J Am Coll Cardiol*. 1997;30(4):878-80.
139. **Kingsnorth AN.** Surgery for periampullary and pancreatic carcinoma: a Liverpool experience. *Ann R Coll Surg Engl*. 1997;79(4):259-63.
140. **Kirshner DL, O'Brien MS, Ricotta JJ.** Risk factors in a community experience with carotid endarterectomy. *J Vasc Surg*. 1989;10(2):178-86.
141. **Kitahata MM, Koepsell TD, Deyo RA, Maxwell CL, Dodge WT, Wagner EH.** Physicians' experience with the acquired immunodeficiency syndrome as a factor in patients' survival. *N Engl J Med*. 1996;334(11):701-6.
142. **Klein LW, Schaer GL, Calvin JE, et al.** Does low individual operator coronary interventional procedural volume correlate with worse institutional procedural outcome?. *J Am Coll Cardiol*. 1997;30(4):870-7.
143. **Konvolinka CW, Copes WS, Sacco WJ.** Institution and per-surgeon volume versus survival outcome in Pennsylvania's trauma centers. *Am J Surg*. 1995;170(4):333-40.
144. **Krakauer H, Shekar SS, Kaye MP.** The relationship of clinical outcomes to status as a Medicare-approved heart transplant center. *Transplantation*. 1995;59(6):840-6.

145. **Kramer S, Meadows AT, Pastore G, Jarrett P, Bruce D.** Influence of place of treatment on diagnosis, treatment, and survival in three pediatric solid tumors. *Journal of Clinical Oncology*. 1984;2(8):917-23.
146. **Kreder HJ, Deyo RA, Koepsell T, Swiontkowski MF, Kreuter W.** Relationship between the volume of total hip replacements performed by providers and the rates of postoperative complications in the state of Washington. *Journal of Bone and Joint Surgery. American Volume*. 1997;79(4):485-94.
147. **Kreder HJ, Williams JI, Jaglal S, Hu R, Axcell T, Stephen D.** Are complication rates for elective primary total hip arthroplasty in Ontario related to surgeon and hospital volumes? A preliminary investigation. *Can J Surg*. 1998;41(6):431-7.
148. **Kuhn EM, Hartz AJ, Krakauer H, Bailey RC, Rimm AA.** The relationship of hospital ownership and teaching status to 30- and 180-day adjusted mortality rates. *Medical Care*. 1994;32(11):1098-108.
149. **Laffel GL, Barnett AI, Finkelstein S, Kaye MP.** The relation between experience and outcome in heart transplantation. *N Engl J Med*. 1992;327(17):1220-5.
150. **Lavernia CJ, Guzman JF.** Relationship of surgical volume to short-term mortality, morbidity, and hospital charges in arthroplasty. *Journal of Arthroplasty*. 1995;10(2):133-40.
151. **Leape LL, Park RE, Solomon DH, Chassin MR, Kosecoff J, Brook RH.** Relation between surgeons' practice volumes and geographic variation in the rate of carotid endarterectomy. *N Engl J Med*. 1989;321(10):653-7.
152. **LeFevre M, Sanner L, Anderson S, Tsutakawa R.** The relationship between neonatal mortality and hospital level. *Journal of Family Practice*. 1992;35(3):259-64.
153. **Lewis SJ.** Trends in cardiac care: contractions, concentration, collaboration. *Hosp Technol Ser*. 1993;12(15):1-29.
154. **Lieberman MD, Kilburn H, Lindsey M, Brennan MF.** Relation of perioperative deaths to hospital volume among patients undergoing pancreatic resection for malignancy. *Annals of Surgery*. 1995;222(5):638-45.
155. **Luft HS, Bunker JP, Enthoven AC.** Should operations be regionalized? The empirical relation between surgical volume and mortality. *N Engl J Med*. 1979;301(25):1364-9.
156. **Luft HS.** The relation between surgical volume and mortality: an exploration of causal factors and alternative models. *Med Care*. 1980;18(9):940-59.
157. **Luft HS, Hunt SS, Maerki SC.** The volume-outcome relationship: practice-makes-perfect or selective-referral patterns? *Health Services Research*. 1987;22(2):157-82.
158. **Luft H, Garnick D, Mark D, McPhee S.** *Hospital Volume, Physician Volume, and Patient Outcomes*; 1990.
159. **Luft HS, Parker JD.** Volume and mortality in coronary artery bypass grafting [published erratum appears in BMJ 1995 Dec 2;311(7018):1508]. *BMJ*. 1995;311(7015):1304-5.
160. **Luft HS.** Pushing good research to go farther. *Health Serv Res*. 1998;33(5 Pt 1):1185-90.
161. **Lupon J, Valle V, Marrugat J, et al.** Six-month outcome in unstable angina patients without previous myocardial infarction according to the use of tertiary cardiologic resources. RESCATE Investigators. Recursos Empleados en el Sindrome Coronario Agudo y Tiempos de Espera. *J Am Coll Cardiol*. 1999;34(7):1947-53.
162. **Maerki SC, Luft HS, Hunt SS.** Selecting categories of patients for regionalization. Implications of the relationship between volume and outcome. *Medical Care*. 1986;24(2):148-58.
163. **Malenka DJ, McGrath PD, Wennberg DE, et al.** The relationship between operator volume and outcomes after percutaneous coronary interventions in high volume hospitals in 1994-1996: the northern New England experience. Northern New England Cardiovascular Disease Study Group. *J Am Coll Cardiol*. 1999;34(5):1471-80.
164. **Manheim LM, Sohn MW, Feinglass J, Ujiki M, Parker MA, Pearce WH.** Hospital vascular surgery volume and procedure mortality rates in California, 1982-1994. *J Vasc Surg*. 1998;28(1):45-56; discussion 56-8.

165. **Marrugat J, Sanz G, Masia R, et al.** Six-month outcome in patients with myocardial infarction initially admitted to tertiary and nontertiary hospitals. RESCATE Investigators. Recursos Empleados en el Sindrome Coronario Agudo y Tiempos de Espera. *J Am Coll Cardiol*. 1997;30(5):1187-92.
166. **Marshall MN, Shekelle PG, Leatherman S, Brook RH.** The public release of performance data: what do we expect to gain? A review of the evidence [In Process Citation]. *JAMA*. 2000;283(14):1866-74.
167. **Mayfield JA, Rosenblatt RA, Baldwin LM, Chu J, Logerfo JP.** The relation of obstetrical volume and nursery level to perinatal mortality. *American Journal of Public Health*. 1990;80(7):819-23.
168. **Maynard C, Every NR, Chapko MK, Ritchie JL.** Institutional volumes and coronary angioplasty outcomes before and after the introduction of stenting. *Eff Clin Pract*. 1999;2(3):108-13.
169. **McArdle CS, Hole D.** Impact of variability among surgeons on postoperative morbidity and mortality and ultimate survival. *BMJ*. 1991;302(6791):1501-5.
170. **McGrath PD, Wennberg DE, Malenka DJ, et al.** Operator volume and outcomes in 12,998 percutaneous coronary interventions. Northern New England Cardiovascular Disease Study Group. *J Am Coll Cardiol*. 1998;31(3):570-6.
171. **McMillan RW, Uppot R, Zibari GB, Aultman DF, Dies DF, McDonald JC.** Can low volume liver transplant centers be successful? The Regional Transplant Center of Willis-Knighton & Louisiana State University Medical Center. The first 120 liver transplants. *J La State Med Soc*. 1999;151(7):367-72.
172. **McWilliam LJ, Knox F, Wilkinson N, Oogarah P.** Performance of skin biopsies by general practitioners. *BMJ*. 1991;303(6811):1177-9.
173. **Meadows AT, Kramer S, Hopson R, Lustbader E, Jarrett P, Evans AE.** Survival in childhood acute lymphocytic leukemia: effect of protocol and place of treatment. *Cancer Investigation*. 1983;1(1):49-55.
174. **Mehta RH, Stalhandske EJ, McCargar PA, Ruane TJ, Eagle KA.** Elderly patients at highest risk with acute myocardial infarction are more frequently transferred from community hospitals to tertiary centers: reality or myth? *Am Heart J*. 1999;138(4 Pt 1):688-95.
175. **Moses LE, Mosteller F.** Institutional differences in Postoperative death rates. Commentary on some of the findings of the National Halothane Study. *JAMA*. 1968;203(7):492-4.
176. **Mullins RJ, Veum-Stone J, Helfand M, et al.** Outcome of hospitalized injured patients after institution of a trauma system in an urban area. *JAMA*. 1994;271(24):1919-24.
177. **Munoz E, Boiardo R, Mulloy K, et al.** Economies of scale, physician volume for orthopedic surgical patients, and the DRG prospective payment system. *Orthopedics*. 1990;13(1):39-44.
178. **Munoz E, Mulloy K, Goldstein J, Tenenbaum N, Wise L.** Costs, quality, and the volume of surgical oncology procedures. *Arch Surg*. 1990;125(3):360-3.
179. **Neoptolemos JP, Russell RC, Bramhall S, Theis B.** Low mortality following resection for pancreatic and periampullary tumours in 1026 patients: UK survey of specialist pancreatic units. UK Pancreatic Cancer Group. *British Journal of Surgery*. 1997;84(10):1370-6.
180. **Nguyen HN, Averette HE, Hoskins W, Penalver M, Sevin BU, Steren A.** National survey of ovarian carcinoma. Part V. The impact of physician's specialty on patients' survival. *Cancer*. 1993;72(12):3663-70.
181. **Norton EC, Garfinkel SA, McQuay LJ, et al.** The effect of hospital volume on the in-hospital complication rate in knee replacement patients.. 1998.
182. **O'Keefe GE, Jurkovich GJ, Copass M, Maier RV.** Ten-year trend in survival and resource utilization at a level I trauma center. *Ann Surg*. 1999;229(3):409-15.
183. **Paneth N, Kiely JL, Wallenstein S, Marcus M, Pakter J, Susser M.** Newborn intensive care and neonatal mortality in low-birth-weight infants: a population study. *New England Journal of Medicine*. 1982;307(3):149-55.
184. **Park RE, Brook RH, Kosecoff J, et al.** Explaining variations in hospital death rates. Randomness, severity of illness, quality of care. *JAMA*. 1990;264(4):484-90.

185. **Parry JM, Collins S, Mathers J, Scott NA, Woodman CB.** Influence of volume of work on the outcome of treatment for patients with colorectal cancer. *Br J Surg*. 1999;86(4):475-81.
186. **Parsonnet V, Bernstein AD.** Outcome versus volume in coronary-bypass operations. *Ann Thorac Surg*. 1996;61(6):1879-80.
187. **Patti M, Corvera CU, Glasgow RE, Way LW.** A hospital's annual rate of esophagectomy influences the operative mortality rate. *Journal of Gastrointestinal Surgery*. 1998;2:186-92.
188. **Pearce WH, Parker MA, Feinglass J, Ujiki M, Manheim LM.** The importance of surgeon volume and training in outcomes for vascular surgical procedures. *J Vasc Surg*. 1999;29(5):768-76; discussion 777-8.
189. **Pellegrini CA, Heck CF, Raper S, Way LW.** An analysis of the reduced morbidity and mortality rates after pancreaticoduodenectomy. *Arch Surg*. 1989;124(7):778-81.
190. **Perler BA, Dardik A, Burleyson GP, Gordon TA, Williams GM.** Influence of age and hospital volume on the results of carotid endarterectomy: a statewide analysis of 9918 cases. *Journal of Vascular Surgery*. 1998;27(1):25-31; discussion 31-3.
191. **Pession A, Locatelli F, Rondelli R, Prete A, Paolucci G.** Current use and outcome of blood and marrow transplantation in childhood according to the Italian Registry. *Bone Marrow Transplant*. 1998;22 Suppl 5:S25-7.
192. **Pett SB, Jr.** Coronary artery bypass grafting volume and mortality. *Ann Thorac Surg*. 1996;62(3):941-3.
193. **Phibbs CS, Bronstein JM, Buxton E, Phibbs RH.** The effects of patient volume and level of care at the hospital of birth on neonatal mortality. *JAMA*. 1996;276(13):1054-9.
194. **Phillips KA, Luft HS, Ritchie JL.** The association of hospital volumes of percutaneous transluminal coronary angioplasty with adverse outcomes, length of stay, and charges in California. *Med Care*. 1995;33(5):502-14.
195. **Phillips KA, Luft HS, Ritchie JL.** Coronary angioplasty procedure volume and major complications. *JAMA*. 1996;275(8):595; discussion 596.
196. **Phillips KA, Luft HS.** The policy implications of using hospital and physician volumes as "indicators" of quality of care in a changing health care environment. *Int J Qual Health Care*. 1997;9(5):341-8.
197. **Pilcher DB, Davis JH, Ashikaga T, et al.** Treatment of abdominal aortic aneurysm in an entire state over 7 1/2 years. *American Journal of Surgery*. 1980;139(4):487-94.
198. **Pollack MM, Alexander SR, Clarke N, Ruttimann UE, Tesselaar HM, Bachulis AC.** Improved outcomes from tertiary center pediatric intensive care: a statewide comparison of tertiary and nontertiary care facilities. *Critical Care Medicine*. 1991;19(2):150-9.
199. **Porter GA, Soskolne CL, Yakimets WW, Newman SC.** Surgeon-related factors and outcome in rectal cancer. *Ann Surg*. 1998;227(2):157-67.
200. **Potosky AL, Warren JL.** Radical prostatectomy: does higher volume lead to better quality? *J Natl Cancer Inst*. 1999;91(22):1906-7.
201. **Pronovost PJ, Jenckes MW, Dorman T, et al.** Organizational characteristics of intensive care units related to outcomes of abdominal aortic surgery. *JAMA*. 1999;281(14):1310-7.
202. **Reis GJ.** Coronary angioplasty procedure volume and major complications. *JAMA*. 1996;275(8):595-6.
203. **Richardson JD, Main KA.** Carotid endarterectomy in the elderly population: a statewide experience. *J Vasc Surg*. 1989;9(1):65-73.
204. **Riley G, Lubitz J.** Outcomes of surgery among the Medicare aged: surgical volume and mortality. *Health Care Financing Review*. 1985;7(1):37-47.
205. **Ritchie JL, Phillips KA, Luft HS.** Coronary angioplasty. Statewide experience in California. *Circulation*. 1993;88(6):2735-43.
206. **Ritchie JL, Maynard C, Chapko MK, Every NR, Martin DC.** Association between percutaneous transluminal coronary angioplasty volumes and outcomes in the Healthcare Cost and Utilization Project 1993-1994. *Am J Cardiol*. 1999;83(4):493-7.

207. **Rogers FB, Osler TM, Shackford SR, Cohen M, Camp L.** Financial outcome of treating trauma in a rural environment. *J Trauma*. 1997;43(1):65-72; discussion 72-3.
208. **Rohrer JE, Vaughan MS, Wenzel RP.** Regionalization of tertiary care: impact of safe cardiovascular volumes in Iowa. *Health Serv Manage Res*. 1997;10(1):1-6.
209. **Romano PS, Mark DH.** Patient and hospital characteristics related to in-hospital mortality after lung cancer resection. *Chest*. 1992;101(5):1332-7.
210. **Roohan PJ, Bickell NA, Baptiste MS, Therriault GD, Ferrara EP, Siu AL.** Hospital volume differences and five-year survival from breast cancer. *American Journal of Public Health*. 1998;88(3):454-7.
211. **Roos LL, Jr., Cageorge SM, Roos NP, Danzinger R.** Centralization, certification, and monitoring. Readmissions and complications after surgery. *Medical Care*. 1986;24(11):1044-66.
212. **Roos LL, Roos NP, Sharp SM.** Monitoring adverse outcomes of surgery using administrative data. *Health Care Financing Review*. 1987;Spec No:5-16.
213. **Roos LL, Sharp SM.** Innovation, centralization, and growth. Coronary artery bypass graft surgery in Manitoba. *Med Care*. 1989;27(5):441-52.
214. **Rosenthal GE, Harper DL, Quinn LM, Cooper GS.** Severity-adjusted mortality and length of stay in teaching and nonteaching hospitals. Results of a regional study. *JAMA*. 1997;278(6):485-90.
215. **Ruby ST, Robinson D, Lynch JT, Mark H.** Outcome analysis of carotid endarterectomy in Connecticut: the impact of volume and specialty. *Ann Vasc Surg*. 1996;10(1):22-6.
216. **Rutledge R, Fakhry SM, Meyer A, Sheldon GF, Baker CC.** An analysis of the association of trauma centers with per capita hospitalizations and death rates from injury. *Ann Surg*. 1993;218(4):512-21; discussion 521-4.
217. **Rutledge R, Oller DW, Meyer AA, Johnson GJ, Jr.** A statewide, population-based time-series analysis of the outcome of ruptured abdominal aortic aneurysm. *Ann Surg*. 1996;223(5):492-502; discussion 503-5.
218. **Ryan TJ.** The critical question of procedure volume minimums for coronary angioplasty. *JAMA*. 1995;274(14):1169-70.
219. **Sagar PM, Hartley MN, MacFie J, Taylor BA, Copeland GP.** Comparison of individual surgeon's performance. Risk-adjusted analysis with POSSUM scoring system. *Dis Colon Rectum*. 1996;39(6):654-8.
220. **Saigal S, Rosenbaum P, Stoskopf B, Sinclair JC.** Outcome in infants 501 to 1000 gm birth weight delivered to residents of the McMaster Health Region. *Journal of Pediatrics*. 1984;105(6):969-76.
221. **Sainsbury R, Haward B, Rider L, Johnston C, Round C.** Influence of clinician workload and patterns of treatment on survival from breast cancer. *Lancet*. 1995;345(8960):1265-70.
222. **Sampalis JS, Lavoie A, Boukas S, et al.** Trauma center designation: initial impact on trauma-related mortality. *J Trauma*. 1995;39(2):232-7; discussion 237-9.
223. **Sampalis JS, Denis R, Frechette P, Brown R, Fleiszer D, Mulder D.** Direct transport to tertiary trauma centers versus transfer from lower level facilities: impact on mortality and morbidity among patients with major trauma. *Journal of Trauma*. 1997;43(2):288-95; discussion 295-6.
224. **Sandrick K.** Learning from experience. In AIDS treatment, knowledge means quality. *Hospitals*. 1993;67(7):32-5.
225. **Schein OD, Steinberg EP, Javitt JC, et al.** Variation in cataract surgery practice and clinical outcomes. *Ophthalmology*. 1994;101(6):1142-52.
226. **Schlenker RE, Hittle DF, Hrincevich CA, Kaehny MM.** Volume/outcome relationships in small rural hospitals. *J Rural Health*. 1996;12(5):395-409.
227. **Schneider EC, Epstein AM.** Influence of cardiac-surgery performance reports on referral practices and access to care. A survey of cardiovascular specialists. *N Engl J Med*. 1996;335(4):251-6.
228. **Schneider EC, Epstein AM.** Use of public performance reports: a survey of patients undergoing cardiac surgery. *JAMA*. 1998;279(20):1638-42.
229. **Segal HE, Rummel L, Wu B.** The utility of PRO data on surgical volume: the example of carotid endarterectomy. *QRB Qual Rev Bull*. 1993;19(5):152-7.

230. **Selby JV, Fireman BH, Lundstrom RJ, et al.** Variation among hospitals in coronary-angiography practices and outcomes after myocardial infarction in a large health maintenance organization. *N Engl J Med*. 1996;335(25):1888-96.
231. **Shapiro MF, Greenfield S.** Experience and outcomes in AIDS. *JAMA*. 1992;268(19):2698-9.
232. **Sheikh K.** The relation between volume and outcome in health care. *N Engl J Med*. 1999;341(14):1085-6.
233. **Shook TL, Sun GW, Burstein S, Eisenhauer AC, Matthews RV.** Comparison of percutaneous transluminal coronary angioplasty outcome and hospital costs for low-volume and high-volume operators. *Am J Cardiol*. 1996;77(5):331-6.
234. **Shortell SM, LoGerfo JP.** Hospital medical staff organization and quality of care: results for myocardial infarction and appendectomy. *Med Care*. 1981;19(10):1041-55.
235. **Shortell SM, Zimmerman JE, Rousseau DM, et al.** The performance of intensive care units: does good management make a difference? *Med Care*. 1994;32(5):508-25.
236. **Showstack JA, Rosenfeld KE, Garnick DW, Luft HS, Schaffarzick RW, Fowles J.** Association of volume with outcome of coronary artery bypass graft surgery. Scheduled vs nonscheduled operations. *JAMA*. 1987;257(6):785-9.
237. **Shroyer AL, Marshall G, Warner BA, et al.** No continuous relationship between Veterans Affairs hospital coronary artery bypass grafting surgical volume and operative mortality. *Annals of Thoracic Surgery*. 1996;61(1):17-20.
238. **Simunovic M, To T, Theriault M, Langer B.** Relation between hospital surgical volume and outcome for pancreatic resection for neoplasm in a publicly funded health care system. *CMAJ*. 1999;160(5):643-8.
239. **Sloan FA, Shayne MW, Doyle MD.** Is there a rationale for regionalizing organ transplantation services? *Journal of Health Politics, Policy and Law*. 1989;14(1):115-67.
240. **Sloan EP, Callahan EP, Duda J, Sheaff CM, Robin AP, Barrett JA.** The effect of urban trauma system hospital bypass on prehospital transport times and Level 1 trauma patient survival. *Ann Emerg Med*. 1989;18(11):1146-50.
241. **Smith DB, Larsson JL.** The impact of learning on cost: the case of heart transplantation. *Hosp Health Serv Adm*. 1989;34(1):85-97.
242. **Sollano JA, Gelijns AC, Moskowitz AJ, et al.** Volume-outcome relationships in cardiovascular operations: New York State, 1990-1995. *J Thorac Cardiovasc Surg*. 1999;117(3):419-28; discussion 428-30.
243. **Solomon RA, Mayer SA, Tarmey JJ.** Relationship between the volume of craniotomies for cerebral aneurysm performed at New York state hospitals and in-hospital mortality. *Stroke*. 1996;27(1):13-7.
244. **Sosa JA, Bowman HM, Gordon TA, et al.** Importance of hospital volume in the overall management of pancreatic cancer. *Annals of Surgery*. 1998a;228(3):429-38.
245. **Sosa JA, Bowman HM, Tielsch JM, Powe NR, Gordon TA, Udelsman R.** The importance of surgeon experience for clinical and economic outcomes from thyroidectomy. *Ann Surg*. 1998b;228(3):320-30.
246. **Sosa JA, Powe NR, Levine MA, Udelsman R, Zeiger MA.** Profile of a clinical practice: Thresholds for surgery and surgical outcomes for patients with primary hyperparathyroidism: a national survey of endocrine surgeons. *J Clin Endocrinol Metab*. 1998c;83(8):2658-65.
247. **Sowden AJ, Deeks JJ, Sheldon TA.** Volume and outcome in coronary artery bypass graft surgery: true association or artefact?. *BMJ*. 1995;311(6998):151-5.
248. **Stanford Center for Health Care Research.** Comparison of hospitals with regard to outcomes of surgery. *Health Serv Res*. 1976;11(2):112-27.
249. **Steele RJ.** The influence of surgeon case volume on outcome in site-specific cancer surgery. *Eur J Surg Oncol*. 1996;22(3):211-3.
250. **Stiller CA.** Centralisation of treatment and survival rates for cancer. *Archives of Disease in Childhood*. 1988;63(1):23-30.

251. **Stiller CA, Draper GJ.** Treatment centre size, entry to trials, and survival in acute lymphoblastic leukaemia. *Archives of Disease in Childhood*. 1989;64(5):657-61.
252. **Stiller CA.** Centralised treatment, entry to trials and survival. *Br J Cancer*. 1994;70(2):352-62.
253. **Stone VE, Seage GRd, Hertz T, Epstein AM.** The relation between hospital experience and mortality for patients with AIDS. *JAMA*. 1992;268(19):2655-61.
254. **Taylor HD, Dennis DA, Crane HS.** Relationship between mortality rates and hospital patient volume for Medicare patients undergoing major orthopaedic surgery of the hip, knee, spine, and femur. *Journal of Arthroplasty*. 1997;12(3):235-42.
255. **Taylor CL, Yuan Z, Selman WR, Ratcheson RA, Rimm AA.** Mortality rates, hospital length of stay, and the cost of treating subarachnoid hemorrhage in older patients: institutional and geographical differences. *Journal of Neurosurgery*. 1997;86(4):583-8.
256. **Tepas JJ, 3rd, Patel JC, DiScala C, Wears RL, Veldenz HC.** Relationship of trauma patient volume to outcome experience: can a relationship be defined? *Journal of Trauma*. 1998;44(5):827-30; discussion 830-1.
257. **Thiemann DR, Coresh J, Oetgen WJ, Powe NR.** The association between hospital volume and survival after acute myocardial infarction in elderly patients. *N Engl J Med*. 1999;340(21):1640-8.
258. **Thorpe AC, Cleary R, Coles J, Vernon S, Reynolds J, Neal DE.** Deaths and complications following prostatectomy in 1400 men in the northern region of England. Northern Regional Prostate Audit Group. *British Journal of Urology*. 1994;74(5):559-65.
259. **Tu JV, Naylor CD.** Coronary artery bypass mortality rates in Ontario. A Canadian approach to quality assurance in cardiac surgery. Steering Committee of the Provincial Adult Cardiac Care Network of Ontario. *Circulation*. 1996;94(10):2429-33.
260. **Tu JV, Hannan EL, Anderson GM, et al.** The fall and rise of carotid endarterectomy in the United States and Canada. *N Engl J Med*. 1998;339(20):1441-7.
261. **Turner BJ, Ball JK.** Variations in inpatient mortality for AIDS in a national sample of hospitals. *J Acquir Immune Defic Syndr*. 1992;5(10):978-87.
262. **Turner BJ, McKee L, Fanning T, Markson LE.** AIDS specialist versus generalist ambulatory care for advanced HIV infection and impact on hospital use. *Med Care*. 1994;32(9):902-16.
263. **Veith FJ, Goldsmith J, Leather RP, Hannan EL.** The need for quality assurance in vascular surgery. *J Vasc Surg*. 1991;13(4):523-6.
264. **Verloove-Vanhorick SP, Verwey RA, Ebeling MC, Brand R, Ruys JH.** Mortality in very preterm and very low birth weight infants according to place of birth and level of care: results of a national collaborative survey of preterm and very low birth weight infants in The Netherlands. *Pediatrics*. 1988;81(3):404-11.
265. **Wachter RM, Luce JM, Turner J, Volberding P, Hopewell PC.** Intensive care of patients with the acquired immunodeficiency syndrome. Outcome and changing patterns of utilization. *Am Rev Respir Dis*. 1986;134(5):891-6.
266. **Wade TP, Halaby IA, Stapleton DR, Virgo KS, Johnson FE.** Population-based analysis of treatment of pancreatic cancer and Whipple resection: Department of Defense hospitals, 1989-1994. *Surgery*. 1996;120(4):680-5; discussion 686-7.
267. **Wen SW, Simunovic M, Williams JI, Johnston KW, Naylor CD.** Hospital volume, calendar age, and short term outcomes in patients undergoing repair of abdominal aortic aneurysms: the Ontario experience, 1988-92. *Journal of Epidemiology and Community Health*. 1996;50(2):207-13.
268. **Wennberg DE, Lucas FL, Birkmeyer JD, Bredenberg CE, Fisher ES.** Variation in carotid endarterectomy mortality in the Medicare population: trial hospitals, volume, and patient characteristics. *JAMA*. 1998;279(16):1278-81.
269. **Wexler MJ.** More procedures, better quality of care? Is there a case for regionalization of pancreatic resection for neoplasm? *CMAJ*. 1999;160(5):671-3.
270. **Wexner SD, Cohen SM, Ulrich A, Reissman P.** Laparoscopic colorectal surgery--are we being honest with our patients? *Dis Colon Rectum*. 1995;38(7):723-7.

271. **Whittle J, Lin CJ, Lave JR, et al.** Relationship of provider characteristics to outcomes, process, and costs of care for community-acquired pneumonia. *Medical Care*. 1998;36(7):977-87.
272. **Williams RL.** Measuring the effectiveness of perinatal medical care. *Medical Care*. 1979;17(2):95-110.
273. **Williamson HA, Jr., Hart LG, Pirani MJ, Rosenblatt RA.** Rural hospital inpatient surgical volume: cutting-edge service or operating on the margin? *J Rural Health*. 1994;10(1):16-25.
274. **Wilson JM, Menkhaus P, Gustin BW.** Volume and outcome of coronary artery bypass graft surgery. *JAMA*. 1987;257(18):2434-5.
275. **Witt PD, Wahlen JC, Marsh JL, Grames LM, Pilgram TK.** The effect of surgeon experience on velopharyngeal functional outcome following palatoplasty: is there a learning curve? *Plast Reconstr Surg*. 1998;102(5):1375-84.
276. **Woods JR, Saywell RM, Jr., Nyhuis AW, Jay SJ, Lohrman RG, Halbrook HG.** The learning curve and the cost of heart transplantation. *Health Serv Res*. 1992;27(2):219-38.
277. **Wright SM, Daley J, Peterson ED, Thibault GE.** Outcomes of acute myocardial infarction in the Department of Veterans Affairs: does regionalization of health care work? *Med Care*. 1997;35(2):128-41.
278. **Yao SL, Lu-Yao G.** Population-based study of relationships between hospital volume of prostatectomies, patient outcomes, and length of hospital stay. *J Natl Cancer Inst*. 1999;91(22):1950-6.
279. **Young GJ, Charns MP, Desai K, et al.** Patterns of coordination and clinical outcomes: a study of surgical services. *Health Serv Res*. 1998;33(5 Pt 1):1211-36.
280. **Zelen J, Bilfinger TV, Anagnostopoulos CE.** Coronary artery bypass grafting. The relationship of surgical volume, hospital location, and outcome. *New York State Journal of Medicine*. 1991;91(7):290-2.
281. **Zimmerman JE, Shortell SM, Knaus WA, et al.** Value and cost of teaching hospitals: a prospective, multicenter, inception cohort study. *Critical Care Medicine*. 1993;21(10):1432-42.

APPENDIX D

When and How Should Purchasers Seek to Selectively Refer Patients to High-Quality Hospitals?

R. Adams Dudley, M.D., M.B.A.; Richard Y. Bae, B.A.; Kirsten L. Johansen, M.D.; and Arnold Milstein, M.D., M.P.H.

There has been evidence for years suggesting important variations in quality among doctors,[1–3] hospitals,[4–7] and health plans,[8,9] and significant effort—such as the development of the Health Plan Employer Data and Information Set (HEDIS) and the Consumer Assessment of Health Plans Study (CAHPS)—has been put into measuring some aspects of quality. Despite this, few consumers, purchasers, health plans, or providers have used this information to obtain high quality care or excellent clinical partners.[10–12] On the contrary, the available evidence suggests that, when selecting among health plans, purchasers give price information much more weight than quality information.[11,13]

The absence of initiatives to act on quality information may reflect a variety of factors. The majority of patients, even those who believe quality varies among providers, are convinced that *their* provider is very good.[14,15] These patients may not feel the need to demand quality improvements. Some purchasers lack the expertise or resources to address quality concerns. Other purchasers may believe that U.S. health care is the best in the world and that they need only ensure that their beneficiaries can access the system to guarantee them high quality care. However, recent reports indicate that quality deficiencies in the United States are common and sometimes severe.[16] Thus, the clinical benefits achievable by directing patients to high quality providers may be significant.

In this paper, we address the rationale for and implementation of selective referral. By selective referral we mean the establishment of policies requiring that patients be: 1) informed that they should go to high quality providers, 2) given a referral to a high quality provider, and 3) offered coverage for the cost of using the high quality provider, even if that provider is outside the patient's usual network of care. The question of when and how to selectively refer patients can

From the Departments of Medicine (R.A.D., K.L.J.) and Epidemiology and Biostatistics (R.A.D., K.L.J.) in the School of Medicine and the Institute for Health Policy Studies (R.A.D., R.Y.B.) at the University of California, San Francisco; the Pacific Business Group on Health (A.M.), and William M. Mercer, Inc. (A.M.).

Corresponding author: R. Adams Dudley, MD, MBA, Institute for Health Policy Studies, University of California, San Francisco, 3333 California St, Suite 265, San Francisco, CA 94118. Phone (415) 476-8617, Fax (415) 476-0705. E-mail: adudley@itsa.ucsf.edu.

TABLE 1 On Which Hospital Characteristic Should Selective Referral Be Based? Characteristics that Serve as Proxies for Quality

- Volume
- Teaching Status
- Level of Care (e.g., Level I Trauma Center)
- Designation by Regulatory Body or Professional Group
- Profit Status
- Outside organizations (e.g., *U.S. News and World Report* 100 "best hospitals")
- Participation in Clinical Trials

be asked at several levels. Purchasers can seek to encourage patients to select the best clinicians, hospitals, or health plans. Alternatively, health plans can direct patients to the best providers. For the sake of exposition, and because there are ongoing initiatives at this level,[17] we will focus on purchasers attempting to ensure that their beneficiaries are referred to high quality hospitals for specific conditions (e.g., the Pacific Business Group on Health's program to ensure patients needing coronary bypass go to the best hospitals). Most of the issues we describe, however, are applicable to initiatives at and among all levels of the health system.

DETERMINING WHICH PROVIDERS ARE HIGH QUALITY: DEFINING THE BASIS FOR SELECTIVE REFERRAL

To optimally direct patients to the highest quality providers, one would need statistically stable risk adjusted measures of both technical quality of care (especially clinical outcomes and processes) and patient satisfaction. Unfortunately, such indices of quality are rarely available. The clinical data needed for risk adjustment are not usually collected for claims or administrative purposes and patient satisfaction with individual clinicians and hospitals is rarely measured at all (CAHPS, for instance, is a measure of satisfaction with *health plans*). There is, however, literature suggesting that many variables may be indirect indicators of quality (Table 1). For example, studies have shown correlation between hospitals' outcomes and their volumes,[7,18] teaching status,[19] level of care,[20] profit status,[21,22] designation by regulatory body or professional group,[23] and participation in clinical trials.[24]

Both direct measurement of quality and the use of indirect indicators like volume have drawbacks as indices of quality (and hence, as bases for selective referral). Direct measurement of outcomes is limited, for many conditions, by hospital sample sizes that are too small to generate annual hospital-specific quality indices.[25,26] For example, in California in 1997, the majority of esophageal cancer surgeries (94 of 169) occurred in hospitals doing 3 or fewer procedures per year. Even a decade of data from such hospitals would not generate stable estimates of an individual hospital's mortality rate. On the other hand, for some conditions, reliance on hospital volume would penalize low volume hospitals that are doing well.[27] Hannan et al. found a volume–outcome relationship for coronary artery bypass grafting (CABG) in New York, but some low volume hospitals had risk adjusted mortality rates below the mean for high volume hospitals.[7] For reasons such as these, purchasers should consider carefully the circumstances under which they would be willing to embark on a selective referral initiative.

There are also some difficult operational issues in the use of volume as a basis for referral. An individual hospital's volume will change frequently and can be dramatically affected by merger with another hospital or acquisition by a hospital chain. Coupled with the usual delays in

reporting, purchasers may have difficulty ascertaining a hospital's true current volume. In addition, some hospitals are very closely affiliated (e.g., multiple hospitals whose medical staffs are on the same medical school faculty) and may have the same person or people performing a single procedure at multiple sites. It is not clear whether volumes in such cases should be combined across sites. As more surgical procedures occur on an outpatient basis, hospitalization databases will become less useful for determining the number of procedures performed at a particular institution. Finally, some may consider the definitions of procedures used in research, on which policy must be based, fairly arbitrary. For example, in counting hospital volumes for esophageal cancer surgeries, most researchers use procedure codes for esophagectomy plus a diagnosis of cancer. Hospitals and physicians may argue, however, that esophageal surgeries done for benign disease should count toward total volume.

Between direct measurement of outcomes and using proxies for quality like volume lie the options of measuring compliance with process indicators (or clinical guidelines) or with structural measures (e.g., ensuring the availability of board certified physicians). However, most providers would argue that this approach should be restricted to those aspects of clinical care or structure for which a clear link to outcomes has been established by randomized trials. Even for common, frequently studied conditions like myocardial infarction, there are many clinical process decisions that probably affect outcome but that cannot be made on this basis of evidence from large randomized trials.[28] Requirements that providers be certified or licensed are even more tenuously linked to outcomes than process measures. Thus, focusing quality measurement exclusively on process or structural variables that correlate with outcome would limit the breadth of quality assessment and penalize providers who have identified methods for improving quality that just happen not to have been studied yet.

CRITERIA FOR ACTION

Types of evidence required to make a decision. In Table 2 we list the categories of data that we believe should go into a decision to pursue (or not) selective referral. The most obvious issue is the expected clinical benefits to the patients referred. This can be measured initially by the observed decrease in mortality or complications at higher quality hospitals, but some attempt should be made to consider whether these benefits are outweighed by the clinical risks associated with transfer. In addition, patients may find referral socially isolating if family or friends cannot accompany them. Disruption of long-term clinical relationships with primary care providers is also a risk.

More subtle, but just as important, is the potential for clinical costs and benefits to patients other than those who are selectively referred. Using the volume–outcome relationship for percutaneous coronary angioplasty as an example, if most patients were referred to high volume hos-

TABLE 2 Types of Evidence Required to Make a Decision Regarding Selective Referral

- Clinical Benefits to Patients Transferred
- Clinical and Social Costs to Patients Transferred
- Clinical Benefits to Other Patients at Referral Hospital
- Clinical and Social Costs to Patients who Remain at the Transferring Hospital
- Economic Implications of Changing the Marketplace Structure

pitals, what would be the clinical implications for those patients who, either because they were too sick to transfer or they did not wish to transfer, remain at an even lower volume hospital? In addition, if the reduction in angioplasty volume makes it impossible for cardiologists to remain at the low volume hospital, what happens to patients with other cardiac problems who lose access to specialty care?

On the other hand, patients who were at high volume hospitals before selective referral may benefit from the increased volume (if practice makes perfect) or may find that additional services—perhaps a cardiac rehabilitation program—are available given the revenues and economies of scale associated with higher volume. Alternatively, though less likely, the new patients may overburden the referral hospital and quality may decline.

Purchasers should also evaluate the potential effects of selective referral on marketplace structure and competition. Selective referral initiatives will increase the market power of preferred hospitals and, if many patients are involved, may lead to the closure of other hospitals. This could allow preferred hospitals to raise prices. In addition, if a hospital characteristic such as volume or teaching status becomes the basis for referral, it will create barriers to entry of new competitors (who must either start at a high volume or immediately meet all criteria necessary to achieve teaching status).

Level of evidence needed to make a decision. Before considering selective referral, purchasers should decide what degree of statistical certainty they will want and how much clinical benefit they will require to justify selective referral. On the first issue, academic statisticians traditionally prefer 95% confidence before concluding that a particular variable (administration of a new drug, or, in this case, a policy initiative) truly affects outcome. However, selective referral could be construed as a social policy, and the criteria for making social decisions vary depending on the perceived consequences of failing to make the right decision. For instance, in American criminal courts, conviction of a defendant should only occur when guilt is established "beyond a reasonable doubt" (presumably a standard greater than 95% confidence). In American civil courts, however, rulings are made against the defendant if the jury finds the defendant "more likely than not" to be guilty (a standard requiring only approximately 51% confidence). Thus, in social decisions in the United States, there is substantial variation in the proof required to initiate an action. Purchasers, not bound by academic tradition, may very well prefer some standard other than 95% confidence, but public discourse on this topic has been essentially nonexistent.

A related issue is the scientific characteristics of the studies on which the decision to selectively refer is based. Specialty societies and consensus panels, when making recommendations about the adoption of technologies or establishing clinical guidelines, frequently comment on the "strength" of the evidence supporting a particular recommendation. These grades traditionally reflect primarily study design (with randomized interventional trials given greater weight than nonrandomized interventional trials or observational studies) and sample size, but may include other considerations such as recruitment bias and loss to follow-up.[29,30] Similar considerations are relevant for instituting selective referral, which is essentially a change in clinical process. However, since most studies of institutional quality are observational, it will be especially important to identify potential sources of bias, particularly with respect to referral of sicker patients to one type of institution or another.

In terms of the level of clinical benefit required to institute selective referral, purchasers should consider whether mortality is the only outcome of interest. For conditions with low overall mortality, differences in mortality may be difficult to demonstrate. However, common com-

TABLE 3 Practical and Political Barriers to Selective Referral

- Access to Preferred Hospital in Rural Areas
- Accommodating Patient Preferences for Care Close to Home
- Ensuring Safe Patient Transfer
- Assuring Capacity at Preferred Hospitals
- Provider Acceptance of Proxies for Risk Adjusted Outcome Measures of Quality
- Health Plan Opposition
- Possible Financial Dissolution of Hospitals

plications or processes of care could be easier to evaluate. To continue with the angioplasty example, many investigators have been unable to identify a statistically significant relationship between volume and mortality, but do find that the rate of "death or emergency CABG or myocardial infarction" falls with increased volume.[31]

It may be that these two issues must be considered together. That is, if there appears to be a mortality benefit, the required level of statistical confidence may be lower than for outcomes of less clinical significance.

POTENTIAL BARRIERS TO SELECTIVE REFERRAL

Table 3 lists both practical and political barriers to the implementation of selective referral. From a practical standpoint, access to preferred hospitals may be difficult, particularly for patients in rural areas. Patient preferences to stay near home, even if this involves using a lower quality hospital, should remain paramount. This necessitates some reporting mechanism so that health plans and medical groups bound by selective referral policies are not penalized by patient decisions. Clinical issues that must be addressed prior to initiation of selective referral include the determination that patients can be transferred safely and that preferred hospitals actually have the capacity to accept the new patients.

Those doctors and hospitals who might lose patients under selective referral are likely to lobby against such initiatives. Their arguments will probably be both technical (especially if an indirect indicator of quality, such as volume, is used) and clinical, focused on the contributions of the less preferred hospitals in other areas that may become collateral damage of a selective referral initiative (e.g., the need to close an emergency room because of loss of profitable cardiac patients). Health plans may also fight selective referral because it involves another reporting requirement and limits their freedom to negotiate with all hospitals in a market.

ALTERNATIVE STRATEGIES TO IMPROVE QUALITY OF CARE

Table 4 lists several options available to purchasers interested in improving the quality of care. Selective referral is just one such approach, and before adopting it purchasers should consider all alternative strategies. We will discuss regulatory options only briefly, because these approaches have mainly fallen out of favor.[27,32]

Regulatory options are not within the power of the purchaser to enact directly, but purchasers could choose to lobby for regulation or legislation. Options targeting health care facilities include mandated regionalization and certificate of need legislation. There have been only a few in-

TABLE 4 Alternative Strategies to Improve Quality of Care

Regulatory Strategies:
- Regulate Targeting Facilities Behavior (e.g., certificate of need, regionalization)
- Regulations Targeting Professional Behavior (e.g., scope of practice limitations)

Competition-Based Strategies (produce quality data to stimulate improvement):
- Consumer-Oriented (e.g., report cards, inclusion in the informed consent process)
- Professional-Oriented (e.g., physician education)
- Purchaser Initiatives (e.g., quality withholds or bonuses, selective referral, direct contracting)

stances of mandated regionalization programs in the United States, though it is more common in Europe. The federal Regional Medical Program of the 1960s was an attempt to mandate regionalization. However, the act offered no structured blueprint for regionalization, instead leaving each region to develop its own plans.[33]

Regionalization was pursued primarily in perinatal care and emergency medical services, especially trauma care. Two controlled studies of perinatal regionalization showed no significant improvement in mortality,[34,35] though one study did show that the areas with regionalization had lower morbidity.[35] Passage of the Emergency Medical Services Act stimulated regionalization of trauma care systems, and early studies suggested reductions in trauma mortality within five years.[36–38] However, these studies used historical controls, making it difficult to determine the extent to which reductions in mortality reflected regionalization vs. secular trends in improved trauma care nationally.[37] A recent cross-sectional analysis controlling for trends in trauma mortality in areas without regionalization suggests that regionalization does, in fact, lower mortality, but the benefits develop only as the system matures over periods as along as a decade.[39]

Certificate of need (CON) legislation was developed as a regulatory mechanism at the state level for review and approval of capital expenditures and service capacity expansions by health care facilities. While CON laws could, in theory, improve access (if used to ensure capacity is well distributed) and prevent waste, they could also limit access (and worsen outcomes) if applied too stringently. In the only study examining the effects of CON laws on hospital mortality rates, states that had more stringent laws had higher risk adjusted mortality rates among Medicare inpatients.[40] However, the risk adjustment methodology used was limited to age, sex, and a small number of comorbid conditions obtained from administrative files, so this finding may not be robust.

Scope of practice laws limit the types of clinical activity in which providers can engage, and hence effectively limit the provision of certain services to specific types of providers. In many states, for instance, optometrists can prescribe corrective lenses, but not medications. This effectively restricts the provision of eye prescriptions to physicians, but few states limit this specifically to ophthalmologists. In theory, one could further limit scope of practice in the name of quality (e.g., if data suggested that board certified ophthalmologists had better outcomes than generalists in caring for patients with glaucoma), but we are aware of no such programs.

While the use of regulation to control quality has declined, programs to use physician, hospital, or health plan performance reports to improve quality have become more common over the last decades, and we focus the remainder of the paper on these. In most cases, proponents of these initiatives expect quality improvements to be mediated through competition.[41] Unfortunately, there have been few evaluations of these initiatives to guide purchasers in the selection of

a strategy. In addition, almost all studies that have been done use historical controls,[41–44] and the few studies that do not are limited by lack of risk adjustment[45] or significant volume differences between types of hospitals.[23] Specific examples will be discussed below, but we first briefly review the positive and negative features of various strategies categorized by target audience.

Consumer-oriented strategies, such as the publication of hospital or health plan report cards, no intervention by the purchaser into clinical processes. However, report cards can be difficult for consumers to understand and, where they have been published so far, appear to have been used by only a small percentage of consumers.[10,46,47]

Approaches that target information to clinicians or medical associations—such as offering educational conferences in which the quality of specific hospitals or types of hospitals are discussed or asking medical associations to certify hospitals—offer the benefit of a greater probability that the recipient audience will understand the data. However, such a strategy will inevitably create conflicts of interest for clinicians at lower quality hospitals, because they risk losing patients by referring them to other institutions. While we believe that altruism can be expected of most clinicians, others will be unwilling to participate.

Inclusion of discussion of the quality of alternative hospitals in the informed consent process would combine informing consumers with clinician-directed approaches. When properly carried out, this tactic is consistent with goals of patient education and empowerment; however, it is also dependent on clinician altruism and vulnerable to the method of presentation of the data by the provider.

Finally, there are other methods in addition to selective referral that focus primarily on purchaser-health plan relationships. For example, purchasers could offer quality withholds or bonuses for health plans that can document excellent risk adjusted outcomes, without requiring the use of preferred hospitals.[48,49] This option allows health plans more contracting and implementation flexibility. However, because it is a less direct attempt to get patients to the best hospitals, it might also mean slower improvement in clinical results.

INFLUENCE OF THE POLICY CONTEXT ON CHOICE OF STRATEGY

Both the choice of a definition of quality—Does one use risk adjusted outcomes? Process measures? Indirect markers like volume?—and the selection of an implementation strategy from among those listed in Table 4 will be heavily influenced by the policy context in which these decisions are made. In particular, the availability of different types of quality data, the impact a quality initiative might have on market structure, and the willingness of various stakeholders to support quality programs are important.

In most states, hospital licensing and discharge databases already exist and can be used to identify teaching or high volume hospitals. Currently, 42 states collect inpatient data, with 30 of those collecting other outpatient data. All states collect at least five discharge diagnoses and procedures; many collect up to 25.[50] However, since hospital outcomes and volume may vary from year to year, policymakers (including purchasers) would be in a better position to categorize hospitals if data from several years were acquired.

For a few conditions in a few states, risk adjusted mortality rates by hospital are available. New York, Pennsylvania, and California, for instance, collect the clinical information necessary to calculate risk adjusted hospital-specific CABG mortality rates. California's system, however, is in its first year, so clear trends in outcome are not yet discernible.

Few states have developed reliable reporting of complications or other morbidity, in part because it is not possible to distinguish, from among the list of discharge diagnoses, conditions that

were present on admission (which should be considered risk factors for bad outcomes) vs. those that developed during the admission (which should be considered complications or bad outcomes). California has just begun to experiment with an indicator that diagnoses were present at admission, but preliminary evaluations by the Office of Statewide Health Planning and Development suggest that coding is inconsistent.

The ability to measure other health outcomes, such as quality of life and time to return to work, would likely increase consumers', purchasers', and policymakers' interest in performance reports. However, to our knowledge, there is no source for such data at this time.

Expected changes in marketplace behavior and competition should also be considered. If one strategy creates an imbalance of market power, policymakers may be more interested in other alternatives. For example, if a proposal to selectively refer patients with a particular condition only to major teaching hospitals would create markets in which plans could only contract with a single hospital (or in which the nearest preferred hospital was very distant), it might be preferable to develop risk adjusted mortality reporting instead—even at substantial administrative cost. In general, the extent to which local clinicians and hospitals are willing to devote resources to quality improvement will also influence one's choice among initiatives that vary in terms of the administrative demands placed on hospitals and plans.

IMPLEMENTATION CONSIDERATIONS

The barriers to implementation vary with the strategy used to introduce quality-based competition, but we will focus on overcoming barriers to selective referral. Since much of the difficulty with selective referral stems from the need to move some patients away from their families and primary providers, programs to encourage selective referral should minimize the impact of distance on patients. If the clinical benefits are significant, it may be worthwhile to pay for family members to accompany patients if that will increase patient compliance with referrals. Similarly, reducing the administrative demands on patients and clinicians associated with changing hospitals may increase acceptance of the program. Finally, it may be prudent to focus initial efforts in areas where success is more likely, such as urban areas with dense concentrations of both patients and hospitals. This will reduce travel distance and facilitate communication between the referral hospitals and primary providers.

In undertaking a program that introduces fundamental changes into the health care market, it will be important to be flexible and to recognize the limitations of the various selective referral strategies as well. For example, for some conditions with low overall volume (e.g., esophageal cancer), it may be impossible to selectively refer based on direct measurement of quality. In this case, selective referral based on minimum volume standards may be optimal. On the other hand, for conditions in which high volume has been shown to improve outcome and for which case loads are large enough to support outcomes measurement (e.g., CABG, in which most hospitals do hundreds of cases a year), it may be possible to offer low volume hospitals a choice: accept selective referral based on volume or agree to participate in direct measurement of outcomes. Matching strategies to the clinical characteristics of different conditions will seem more reasonable to clinicians and hospitals than choosing a universal approach.

EXAMPLES OF THE USE OF QUALITY INFORMATION TO INFLUENCE REFERRAL PATTERNS

Though not yet commonplace, there have been several initiatives to collect and utilize quality information to improve care (Table 5). The organizations sponsoring these projects are primarily state governments, private purchasers, and the Health Care Financing Administration (HCFA), but companies that sell quality data or provide it on the Internet and profit from associated advertising also exist. Most of the programs involve the use of HEDIS, patient satisfaction surveys, or the collection and publication of hospital-specific risk adjusted mortality rates. We will discuss these initiatives in terms of their sponsors, but in Table 5 we sort the programs discussed below by the type of quality information produced, the target audience, and whether the target audience is also given financial incentives to select high quality organizations.

There are several initiatives run by states. Since 1989, the New York State Department of Health has had a Cardiac Surgery Reporting System (CSRS) through which it collects clinical data on all patients undergoing CABG. The Department reports to hospitals both risk adjusted mortality rates and lists of preoperative risk factors that are significantly related to mortality. In addition, risk adjusted mortality rates for each hospital and surgeon that perform CABG are released in public reports. In each of the first four years after CSRS was introduced, actual and risk adjusted mortality decreased. Actual mortality decreased from 3.52% in 1989 to 2.78% to 1992, a 21% reduction. Risk adjusted mortality decreased 41% from 4.17% in 1989 to 2.45% in 1992.[42] It is unknown whether public reporting of quality was the sole or predominant cause of the mortality reduction, as there was a simultaneous nationwide trend toward lower CABG mortality.[51] However, comparison of the reduction in mortality in New York to national trends (done without risk adjustment) did suggest that mortality fell faster in New York.[45] If there is an effect of public reporting, it is also impossible to know how much of this effect comes from reporting hospital data and how much is associated with reporting individual surgeon performance.

Pennsylvania also developed a CABG mortality report in the early 1990s. Risk adjusted mortality dropped from 3.9% to 2.9% (or 26%) between 1990 and 1993.[52] In 1994–1995, the decrease in mortality was accompanied by a decline in charges for CABG of 3.9% from the previous year even though the number of CABG surgeries rose 25% from 1991 through 1995.[53] Interestingly, though intended for consumers, the report has been more useful to hospitals and providers in improving mortality outcomes than to consumers looking for high quality providers. Only 12% of patients undergoing CABG were aware of the report prior to surgery.[10]

California produces annual hospital mortality reports for myocardial infarction (MI) and is developing reports for CABG, pneumonia, maternal outcomes, hip fracture, and intensive care unit (ICU) patients. To date, the MI reports have not yet been used extensively by hospitals[54] or purchasers to improve quality. Missouri has an obstetrics quality public reporting system that includes mortality, morbidity, and patient satisfaction measures. Longo et al. found that the implementation of this system was associated with improvements in all measures, but this study used historical controls.[44]

Private purchasers across the country also have begun to collect quality data. The 1998 National Business Coalition on Health Survey found that many business coalitions gather data on quality of care, incorporate financial incentives for performance into purchase contracts, or collaborate with plans or providers on continuous quality improvement efforts.[49] Having acquired quality data, some coalitions have attempted to stimulate quality-based competition.

TABLE 5 Alternative Strategies to Introduce Quality Information as a Determinant of Market Behavior, with Recent Examples

	Method of Introducing Data to the Market					
	Offer information but not financial incentives, targeted to:			Offer information and financial incentives, targeted to:		
Quality Information Generated	Patients	Physicians, Hospitals, or Health Systems	Health Plans	Patients	Physicians, Hospitals, or Health Systems	Health Plans
Clinical Outcomes—Mortality	CA, Cleve, DFW, HCFA/Mort, HlthGrade, MO, NY, PA, *U.S. News*	CA, DFW, HCFA/Mort, HlthGrade, MO, NY, PA, *U.S. News*	CA, HCFA/Mort, HlthGrade, MO, NY, PA		Cleve, HCFA/CtrsExcel, HCFA/Heart	PBGH/EBHR for CABG
Clinical Outcomes—Morbidity or Complications	CA, Cleve, DFW, HEDIS, HlthGrade, MO, St L	CA, DFW, HEDIS, MO	CA, HEDIS, MO	Digital, GM, GTE	Cleve, HCFA/CtrsExcel, HCFA/Heart	Digital, GM, GTE, St L
Patient Satisfaction	BHCAG, Cleve, Chicago, CAHPS, DFW, HEDIS, MO	CAHPS, DFW, HEDIS, MO	CAHPS, HEDIS, MO	Digital, GM, GTE	Cleve, HCFA/Heart	Digital, GM, GTE

Process Measures	HEDIS, St L	HEDIS	HEDIS	Digital, GM, GTE	HCFA/Heart, HCFA/CtrsExcel	Digital, GM, GTE, St L
Structural Measures	*U.S. News*	*U.S. News*			HCFA/CtrsExcel	
Hospital Volume	PBGH/Hlthscope, HlthGrade				HCFA/CtrsExcel	Leap, GM, PBGH/EBHR for 6 non-CABG procedures

NOTE: (1) In compiling this table, we considered a Caesarean section procedure a complication or morbidity, not a process measure. (2) When a purchaser discounted prices to employees of higher quality plans, this was considered a financial incentive to both patients and plans.

LEGEND: BHCAG = Buyers Health Care Action Group; CA = California Hospital Outcomes Project; CABG = coronary artery bypass grafting; Chicago = Chicago Business Group on Health; Cleve = Cleveland Health Quality Choice; CAHPS = Consumer Assessment of Health Plans Study; DFW = Dallas/Fort Worth Business Group on Health; Digital = Digital Equipment Corp. health benefits plan; GM = General Motors health benefits plan; GTE = GTE, Inc., health benefits plan; HCFA/CtrsExcel = Health Care Financing Administration Centers of Excellence Program; HCFA/Heart = Health Care Financing Administration heart transplantation approval process; HCFA/Mort = Health Care Financing Administration hospital mortality reports; HEDIS = Health Plan Employer Data and Information Set 3.0; HlthGrade = HealthGrades.com web site; Leap = The Leapfrog Group; MO = Missouri Obstetrics Reporting Program; NY = New York State Cardiac Surgery Reporting System; PBGH/Hlthscope = Pacific Business Group on Health's Healthscope web site; PBGH/EBHR = Pacific Business Group on Health Evidence-Based Hospital Referral Program; PA = Pennsylvania Health Care Cost Containment Council; St L = Gateway Purchasing Association; *U.S. News* = *U.S. News and World Report's* Top 100 Hospitals.

One of the oldest such programs was Cleveland Health Quality Choice (CHQC). In 1989, a group of large employers in the Cleveland area decided to collect hospital quality information and use it in both public reports and direct contracting. Measures collected included risk adjusted hospital mortality for certain conditions, cesarean section rates, and patient satisfaction. As reports came out in the early 1990s, hospital mortality rates declined,[43] but again only historical controls were available. CHQC ended in 1999 when some hospitals, unhappy with the accuracy of the quality data (especially ICU mortality measurement) and the data collection burdens and unconvinced that purchasers were really using the data to direct patients to high quality hospitals, withdrew from the reporting system.[55]

A hospital reporting system similar to CHQC has been developed by the Dallas-Fort Worth Business Group on Health, but this program only reports the data back to hospitals and to consumers, without any use of financial or volume incentives. There are plans to develop a program to measure physician quality, but this has not been implemented.[56]

Since 1996, the Chicago Business Group on Health has provided public reports of patient satisfaction with health plans as a demonstration site for the National Committee on Quality Assurance's satisfaction survey. An evaluation is planned but has not been published.[56] In Minneapolis-St. Paul, the Buyers Health Care Action Group also provides patient satisfaction data to consumers, and care systems that score well seem to gain market share.[57]

The Pacific Business Group on Health (PBGH) has several programs in place. On the PBGH HealthScope web site,[58] PBGH offers consumers a synopsis of the literature on volume–outcome relationships and lists several conditions for which higher volume is associated with lower mortality. The site also allows consumers to look at the condition-specific volume for hospitals in their area.

PBGH is also attempting to encourage evidence-based hospital referral through its health plan contracting process. With the state of California, PBGH has developed risk adjusted CABG mortality reports and is requiring that its contracting plans report which hospitals are used for patients undergoing CABG. In addition, based on recent reports of the potential to lower mortality in California through referral of patients away from low volume hospitals,[18] PBGH is requiring that health plans ensure patients with selected conditions go to high volume hospitals.

Some large employers have acted on their own to stimulate quality. General Motors (GM), GTE, and Digital Equipment Corporation all discount the premiums employees must pay if they choose higher quality plans from among those the company offers (based on HEDIS scores and patient satisfaction surveys).[56] This provides financial incentives both to patients in the form of lower prices and to plans in the form of higher market share. These programs have not been formally evaluated for their impact on quality of care.

The Leapfrog Group, a newly formed organization comprising many large employers (GM, Ford, GTE, GE, and others) and purchasing coalitions (including PBGH), is also pursuing evidence-based hospital referral. This organization has developed a set of health plan performance standards that include volume standards for specific conditions, but is also considering exemptions from volume standards if hospitals agree instead to participate in risk adjusted mortality measurement.

We are aware of only one study of the impact of a selective referral program. In 1986, HCFA created a heart transplantation program for Medicare beneficiaries. This involved a transplant facility approval process that included a volume standard—performing at least 500 cardiac catheterizations and 250 cardiac surgeries (of all types, not just transplants) annually. Other standards involved patient selection and management processes, resources available, reporting re-

quirements and survival rates.[59] In a study of heart transplants done at approved versus nonapproved centers, the probability of death at a Medicare-approved center was 7.0 ± 0.4% at 30 days and 16.2% ± 0.6% at one year post transplantation. For nonapproved centers, the probability of death at 30 days was 9.2 ± 0.4% and 19.2% ± 0.6% at one year ($p < .05$ for both).[23] This suggests that selective referral for specific procedures can improve outcomes if the selection criteria are legitimate measures of quality.

HCFA has had several other programs aimed at improving quality. In the 1980s, HCFA released hospital-specific mortality rates among Medicare patients. Hospital executives were skeptical about the accuracy and usefulness of the HCFA reports,[60] and consumers and purchasers appear to have paid little attention. A study of New York hospitals found no effect of mortality on occupancy rates,[61]while a national study found a statistically significant but very small effect on utilization.[62] HCFA also has a Centers of Excellence program in which hospitals apply to receive this designation for CABG or orthopedic surgery. In considering applications, HCFA uses a volume threshold, but also considers each hospital's outcomes and evaluates institutional processes and structure.[63] At this time, there have been no studies of the impact of this program published in the peer-reviewed literature.

Some companies are discovering commercial uses for health care quality information. HealthGrades has created a web site that evaluates hospitals, physicians, health plans, and HMOs.[64] In ranking hospitals, HealthGrades utilized the HCFA Centers of Excellence demonstration program minimum volume requirements for cardiac surgery and orthopedic surgery. For the other conditions, HealthGrades.com classified the hospitals within the lowest quartile with respect to volume as "lower volume hospitals." HealthGrades then uses HCFA discharge data to calculate condition-specific risk adjusted mortality rates. The data are available to consumers for free, and the site is supported by advertising, only some of which is for health-related products.

U.S. News and World Report has been publishing ratings of hospitals since 1991. The rating process includes mortality measurements, structural measures, and hospital volume.[65] It purports to include process measures as well, but these are based on reputational surveys. We are aware of no evaluation of the impact of these reports on outcomes.

FUTURE DIRECTIONS

Much research is needed to better understand how best to use quality data to improve health care. The two primary needs are to determine how best to use quality information once it is generated[66] and how it is that physicians and hospitals with better outcomes achieve these results.[67]

As a first step in the evaluation of the use of quality information, studies of the many initiatives shown in Table 5 should be performed. Specific questions include whom to target for dissemination of quality data and whether information is enough or financial incentives are necessary. An important issue that needs to be considered is whether all patients should receive the same data, or whether, for example, patients with chronic diseases might be interested in different information than the general population.

In performing these analyses, investigators should attempt to find better controls than the typical "before and after" patient groups. One health plan-oriented approach would be to analyze whether, as the plan implements, for example, selective referral to high volume hospitals for a cancer surgery, the plan-level mortality rate falls faster than the contemporaneous state-wide mortality rate for the same procedure.

Understanding why high performers do well will also be critical. This strategy can be applied whether physicians or hospitals are identified as superior based on direct measurement of quality

or based on indirect markers like volume. Initial steps to do this have focused on whether high and low performers differ in processes that have been shown by randomized controlled trials to improve outcomes, and this is indeed an important question. However, investigators should also consider the possibility that better outcomes are due to differences in process that have not yet been studied in randomized trials. As the determinants of higher quality are identified, it will be important to assess how easily they can be transferred to institutions with lower quality. For instance, if cancer centers with better surveillance procedures for chemotherapy-related bone marrow suppression have lower mortality, these algorithms can easily be shared with other centers (since they involve only blood tests and diligence). If, alternatively, it appears that the cancer centers with fewer marrow suppression-related fatalities are those that have full-time on site board certified hematologists, it may be difficult for other centers to have enough clinical activity to be able to support that level of hematology coverage. Answering questions like these will be necessary to maximize the quality gain for any given investment in health care improvement.

CONCLUSIONS

The clinical impact of efforts to introduce quality as a basis of competition in the marketplace will depend greatly on the ability of purchasers or policymakers to choose a strategy appropriate to a given location and to address the implementational barriers described above. In theory, referral based on actual outcomes and processes or public reporting of quality data are preferable to referral based on indirect indicators of quality like hospital volume. However, selective referral based on volume is currently much easier to achieve. Introducing selective referral based on volume as a first step may create an impetus for developing databases that can foster condition-specific outcome or process measurement. Further research is needed to allow comparison of the costs and benefits of the various strategies listed in Tables 4 and 5.

Regardless of the strategy selected, any quality improvement program should include an evaluation component to determine that the estimates of clinical benefit on which the choice of strategy was based were accurate. In addition, an assessment of which patients were transferred—and which were not and the reasons that patients were not transferred—must be performed to ensure that the program is implemented optimally and the benefits of referral are distributed fairly. Finally, the presence of significant barriers to selective referral and other approaches designed to introduce quality competition imply that the long-term success of such initiatives will depend on ongoing support from purchasers and policymakers.

REFERENCES

1. Burns LR, Wholey DR. The effects of patient, hospital, and physician characteristics on length of stay and mortality. *Med Care.* 1991;29:251–71.
2. Edwards WH, Morris JA, Jr., Jenkins JM, Bass SM, MacKenzie EJ. Evaluating quality, cost-effective health care. Vascular database predicated on hospital discharge abstracts. *Ann Surg.* 1991;213:433–8.
3. Ellis SG, Weintraub W, Holmes D, Shaw R, Block PC, King SB, 3rd. Relation of operator volume and experience to procedural outcome of percutaneous coronary revascularization at hospitals with high interventional volumes. *Circulation.* 1997;95:2479–84.
4. Glasgow RE, Mulvihill SJ. Hospital volume influences outcome in patients undergoing pancreatic resection for cancer. *West J Med.* 1996;165:294–300.
5. Hosenpud JD, Breen TJ, Edwards EB, Daily OP, Hunsicker LG. The effect of transplant center volume on cardiac transplant outcome. A report of the United Network for Organ Sharing Scientific Registry. *JAMA.* 1994;271:1844–9.

6. Patti MG, Corvera CU, Glasgow RE, Way LW. A hospital's annual rate of esophagectomy influences the operative mortality rate. *J Gastrointest Surg*. 1998;2:186–92.
7. Hannan EL, Kilburn H, Jr., Bernard H, O'Donnell JF, Lukacik G, Shields EP. Coronary artery bypass surgery: the relationship between inhospital mortality rate and surgical volume after controlling for clinical risk factors. *Med Care.* 1991;29:1094–107.
8. Miller RH, Luft HS. Managed care plan performance since 1980. A literature analysis. *JAMA.* 1994; 271:1512–9.
9. Chernew M, Scanlon DP. Health plan report cards and insurance choice. *Inquiry.* 1998;35:9–22.
10. Schneider EC, Epstein AM. Use of public performance reports: a survey of patients undergoing cardiac surgery. *JAMA.* 1998;279:1638–42.
11. McLaughlin CG. Health care consumers: choices and constraints. *Med Care Res Rev.* 1999;56:24–59; discussion 60–6.
12. Schneider EC, Epstein AM. Influence of cardiac-surgery performance reports on referral practices and access to care. A survey of cardiovascular specialists. *N Engl J Med.* 1996;335:251–6.
13. Langreth R. Health-Care Costs Rise for Employers, According to Poll. Wall Street Journal. September 8, 1997, B2.
14. Mundinger MO, Kane RL, Lenz ER, et al. Primary care outcomes in patients treated by nurse practitioners or physicians: a randomized trial. JAMA. 2000;283:59–68.
15. Murray-Garcia JL, Selby JV, Schmittdiel J, Grumbach K, Quesenberry CP, Jr. Racial and ethnic differences in a patient survey: patients' values, ratings, and reports regarding physician primary care performance in a large health maintenance organization. *Med Care.* 2000;38:300–10.
16. Kohn LT, Corrigan JM, Donaldson MS. To Err is Human: Building a Safer Health Care System. Washington, D. C.: Institute of Medicine; 1999.
17. PBGH. Two New PBGH Initiatives to Make Provider Quality Count More. Pacific Currents. Fall, 1998, 3; 5.
18. Dudley RA, Johansen KL, Brand R, Rennie DJ, Milstein A. Selective referral to high-volume hospitals: estimating potentially avoidable deaths. *JAMA.* 2000;283:1159–66.
19. Rosenthal GE, Harper DL, Quinn LM, Cooper GS. Severity-adjusted mortality and length of stay in teaching and nonteaching hospitals. Results of a regional study. *JAMA.* 1997;278:485–90.
20. Phibbs CS, Bronstein JM, Buxton E, Phibbs RH. The effects of patient volume and level of care at the hospital of birth on neonatal mortality. *JAMA.* 1996;276:1054–9.
21. Hartz AJ, Krakauer H, Kuhn EM, et al. Hospital characteristics and mortality rates. *N Engl J Med.* 1989;321:1720–5.
22. Kuhn EM, Hartz AJ, Krakauer H, Bailey RC, Rimm AA. The relationship of hospital ownership and teaching status to 30- and 180-day adjusted mortality rates. *Med Care.* 1994;32:1098–108.
23. Krakauer H, Shekar SS, Kaye MP. The relationship of clinical outcomes to status as a Medicare-approved heart transplant center. *Transplantation.* 1995;59:840–6.
24. Wennberg DE, Lucas FL, Birkmeyer JD, Bredenberg CE, Fisher ES. Variation in carotid endarterectomy mortality in the Medicare population: trial hospitals, volume, and patient characteristics. *JAMA.* 1998;279:1278–81.
25. Hofer TP, Hayward RA. Identifying poor-quality hospitals. Can hospital mortality rates detect quality problems for medical diagnoses? *Med Care.* 1996;34:737–53.
26. Zalkind DL, Eastaugh SR. Mortality rates as an indicator of hospital quality. *Hosp Health Serv Adm.* 1997;42:3–15.
27. Phillips KA, Luft HS. The policy implications of using hospital and physician volumes as "indicators" of quality of care in a changing health care environment. Int J Qual Health Care. 1997;9:341–8.
28. Ryan TJ, Antman EM, Brooks NH, et al. 1999 update: ACC/AHA guidelines for the management of patients with acute myocardial infarction. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Management of Acute Myocardial Infarction). *J Am Coll Cardiol.* 1999;34:890–911.

29. Hayward RS, Wilson MC, Tunis SR, Bass EB, Guyatt G. Users' guides to the medical literature. VIII. How to use clinical practice guidelines. A. Are the recommendations valid? The Evidence-Based Medicine Working Group. *JAMA.* 1995;274:570–4.
30. Wilson MC, Hayward RS, Tunis SR, Bass EB, Guyatt G. Users' guides to the Medical Literature. VIII. How to use clinical practice guidelines. B. what are the recommendations and will they help you in caring for your patients? The Evidence-Based Medicine Working Group. *JAMA.* 1995;274: 1630–2.
31. Jollis JG, Peterson ED, Nelson CL, et al. Relationship between physician and hospital coronary angioplasty volume and outcome in elderly patients. *Circulation.* 1997;95:2485–91.
32. Chassin MR. Assessing strategies for quality improvement. *Health Aff (Millwood).* 1997;16:151–61.
33. Marston RQ, Yordy K. A nation starts a program: Regional Medical Programs, 1965–1966. *J Med Educ.* 1967;42:17–27.
34. McCormick MC, Shapiro S, Starfield BH. The regionalization of perinatal services. Summary of the evaluation of a national demonstration program. *JAMA.* 1985;253:799–804.
35. Siegel E, Gillings D, Campbell S, Guild P. A controlled evaluation of rural regional perinatal care: impact on mortality and morbidity. *Am J Public Health.* 1985;75:246–53.
36. Mullins RJ, Veum-Stone J, Hedges JR, et al. Influence of a statewide trauma system on location of hospitalization and outcome of injured patients. *J Trauma.* 1996;40:536–45; discussion 545–6.
37. Mullins RJ, Veum-Stone J, Helfand M, et al. Outcome of hospitalized injured patients after institution of a trauma system in an urban area. *JAMA.* 1994;271:1919–24.
38. Stewart TC, Lane PL, Stefanits T. An evaluation of patient outcomes before and after trauma center designation using Trauma and Injury Severity Score analysis. *J Trauma.* 1995;39:1036–40.
39. Nathens AB, Jurkovich GJ, Cummings P, Rivara FP, Maier RV. The Effect of Organized Systems of Trauma Care on Motor Vehicle Crash Mortality. *JAMA.* 2000;283:1990–1994.
40. Shortell SM, Hughes EF. The effects of regulation, competition, and ownership on mortality rates among hospital inpatients. *N Engl J Med.* 1988;318:1100–7.
41. Marshall MN, Shekelle PG, Leatherman S, Brook RH. The Public Release of Performance Data-What do we Expect to Gain? A Review of the Evidence. *JAMA.* 2000;283:1866–1874.
42. Hannan EL, Kilburn H, Jr., Racz M, Shields E, Chassin MR. Improving the outcomes of coronary artery bypass surgery in New York State. *JAMA.* 1994;271:761–6.
43. Rosenthal GE, Quinn L, Harper DL. Declines in hospital mortality associated with a regional initiative to measure hospital performance. *Am J Med Qual.* 1997;12:103–12.
44. Longo DR, Land G, Schramm W, Fraas J, Hoskins B, Howell V. Consumer reports in health care. Do they make a difference in patient care? *JAMA.* 1997;278:1579–84.
45. Peterson ED, DeLong ER, Jollis JG, Muhlbaier LH, Mark DB. The effects of New York's bypass surgery provider profiling on access to care and patient outcomes in the elderly. *J Am Coll Cardiol.* 1998;32:993–9.
46. Hibbard JH, Jewett JJ. Will quality report cards help consumers? *Health Aff (Millwood).* 1997;16: 218–28.
47. Wicks EK, Meyer JA. Making report cards work. *Health Aff (Millwood).* 1999;18:152–5.
48. Fraser I, McNamara P, Lehman GO, Isaacson S, Moler K. The pursuit of quality by business coalitions: a national survey. *Health Aff (Millwood).* 1999;18:158–65.
49. Schauffler HH, Brown C, Milstein A. Raising the bar: the use of performance guarantees by the Pacific Business Group on Health. *Health Aff (Millwood).* 1999;18:134–42.
50. NAHDO, MEDSTAT. The National Association for Health Data Organizations, The MEDSTAT Group, Statewide Encounter-Level Inpatient and Outpatient Data Collection Activities, Summary Report. Salt Lake City, UT: Agency for Health Care Policy and Research; 1999.
51. Grover FL. The Society of Thoracic Surgeons National Database: current status and future directions. Ann Thorac Surg. 1999;68:367–73.
52. Moore JD, Jr. The public eye. Published outcomes reports effect change at Pa. hospitals. *Mod Healthc.* 1997;27:140, 155.

53. Moore JD, Jr. Real results. Pa. report gives consumers outcomes data they can use. *Mod Healthc.* 1998;28:46.
54. Luce JM, Thiel GD, Holland MR, Swig L, Currin SA, Luft HS. Use of risk-adjusted outcome data for quality improvement by public hospitals. *West J Med.* 1996;164:410–4.
55. Burton TM. Operation that Rated Hospitals Was Success, But the Patience Died. Wall Street Journal. August 23, 1999, A1.
56. Meyer J, Ribowski L, Eichler R. Theory and Reality of Value-Based Purchasing: Lessons from the Pioneers. Rockville, MD: Agency for Health Care Policy and Research; 1997.
57. Christianson J, Feldman R, Weiner JP, Drury P. Early experience with a new model of employer group purchasing in Minnesota. *Health Aff (Millwood).* 1999;18:100–14.
58. PBGH. http://www.HealthScope.org. Accessed April 26, 2000.
59. Medicare program: Proposed criteria for Medicare coverage of Heart Transplants. Fed Register. 1986;51:37164–70.
60. Berwick DM, Wald DL. Hospital leaders' opinions of the HCFA mortality data. *JAMA.* 1990;263: 247–9.
61. Vladeck BC, Goodwin EJ, Myers LP, Sinisi M. Consumers and hospital use: the HCFA "death list." Health Aff (Millwood). 1988;7:122–5.
62. Mennemeyer ST, Morrisey MA, Howard LZ. Death and reputation: how consumers acted upon HCFA mortality information. *Inquiry.* 1997;34:117–28.
63. Wilensky GR. From the Health Care Financing Administration. *JAMA.* 1991;266:3404.
64. HealthGrades.com. http://www.healthgrades.com. Accessed April 27, 2000.
65. Hill CA, Winfrey KL, Rudolph BA. "Best hospitals": a description of the methodology for the Index of Hospital Quality. *Inquiry.* 1997;34:80–90.
66. Jencks SF. Clinical Performance Measurement—A Hard Sell. *JAMA.* 2000;283:2015–2016.
67. Hannan EL. The relation between volume and outcome in health care. *N Engl J Med.* 1999;340: 1677–9.