Interpreting the Volume-Outcome Relationship in the Context of Cancer Care

Maria Hewitt and Diana Petitti, Editors, National Cancer Policy Board, Division on Earth and Life Studies, National Research Council

ISBN: 0-309-51091-0, 42 pages, 6 x 9, (2001)

Interpreting the Volume–Outcome Relationship in the Context of Cancer Care

Maria Hewitt and Diana Petitti**,** *Editors*

National Cancer Policy Board
INSTITUTE OF MEDICINE

and

Division on Earth and Life Studies
NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY PRESS
Washington, D.C.

NATIONAL ACADEMY PRESS • 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 National Cancer Institute; the Centers for Disease Control and Prevention; the American Cancer Society; American Society of Clinical Oncology; Abbott Laboratories; Amgen, Inc.; and Aventis. The views presented in this report are those of the National Cancer Policy Board of the Institute of Medicine and the National Research Council and are not necessarily those of the funding agencies.

International Standard Book Number 0-309-07586-6

Additional copies of this report are available for sale from the National Academy Press, 2101 Constitution Avenue, N.W., Box 285, Washington, D.C. 20055. Call (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area), or visit the NAP's home page at **www.nap.edu.** 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 NATIONAL ACADEMIES

National Academy of Sciences
National Academy of Engineering
Institute of Medicine
National Research Council

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.

NATIONAL CANCER POLICY BOARD

Arnold J. Levine (C*hair*), President, The Rockefeller University, New York

Joseph Simone (*Vice Chair*), Medical Director, Huntsman Cancer Foundation and Institute, University of Utah, Salt Lake City

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

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

Tim Byers, Professor of Epidemiology and Program Leader, Clinical Cancer Prevention and Control, University of Colorado Health Sciences Center, Denver

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

Susan Curry, Professor of Health Policy and Administration and Director, Health Research and Policy Centers, University of Illinois at Chicago

Norman Daniels, Professor of Philosophy, Tufts University, Boston *(member through April 2001)*

Kathleen Foley, Director, Project on Death in America, The Open Society and Memorial Sloan-Kettering Cancer Center, New York *(member through April 2001)*

Thomas Kelly, Professor and Chairman, Department of Molecular Biology and Genetics, The Johns Hopkins University School of Medicine, Baltimore *(member through April 2001)*

Mark McClellan, Assistant Professor of Economics, Stanford University *(member through March 2001)*

William McGuire, Chief Executive Officer, UnitedHealth Group, Minnetonka, MN

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

Monica Morrow, Professor of Surgery and Director, Lynn Sage Comprehensive Breast Program, Northwestern University Medical School, Chicago

Nancy Mueller, Professor of Epidemiology, Harvard University School of Public Health, Boston

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

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

John Seffrin, Chief Executive Officer, American Cancer Society, Atlanta

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

Frances Visco, President, National Breast Cancer Coalition, Washington, D.C. *(member through April 2001)*

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

Study Staff
Maria Hewitt, Study Director
Florence Poillon, Editor

NCPB Staff
Roger Herdman, Director, National Cancer Policy Board
Nicci T. Dowd, Administrator
Jennifer Cangco, Financial Associate

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:

John Cameron, Department of Surgery, The Johns Hopkins University School of Medicine
Jeffrey Crane, Rex Cancer Center
Edward Hannan, Department of Health Policy, Management, and Behavior, State University of New York at Albany
Arthur Levin, Center for Medical Consumers
Arnold Milstein, William M. Mercer, Inc.
Scott Parker, Retired, Intermountain Health Care, Inc.

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, The Primary Care Outcomes Research Institute, New England Medical Center Hospitals, appointed by the Institute of Medicine and Joe Newhouse, Department of Health Policy and Management, Harvard University, appointed by the NRC's Report Review Committee, who were 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 board and the institution.

Contents

White Paper

Interpreting the Volume–Outcome Relationship in the Context of Cancer Care

National Cancer Policy Board

BACKGROUND

A higher volume of care translates into improved short-term outcomes for certain complex treatments for cancer. As much as a threefold increase in deaths following esophagectomy and pancreatectomy in lower- as compared to higher-volume hospitals has, for example, been reported in the health services research literature. These findings prompted the National Cancer Policy Board (board) to recommend in its 1999 report *Ensuring Quality Cancer Care* that cancer care is optimally delivered in systems of care that

> Ensure that patients undergoing procedures that are technically difficult to perform and have been associated with higher mortality in lower-volume settings receive care at facilities with extensive experience (i.e., high-volume facilities). Examples of such procedures include removal of all or part of the esophagus, surgery for pancreatic cancer, removal of pelvic organs, and complex chemotherapy regimens.

Although evidence of the relationship between higher volume and better outcomes is strong and consistent for certain relatively uncommon procedures, the board did not have evidence to support a broader application of its recommendation. Many questions arose in board deliberations regarding the nature of the relationship and the processes of care that might explain it. Furthermore, the board recognized potential difficulties in implementing policies to concentrate care into higher-volume settings and decided that such issues had to be explored further.

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

1. review evidence of the relationship between volume of services and health-related outcomes for cancer and other conditions;
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 structured around presentations of two commissioned papers:

1. "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; and
2. "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.

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 (AHRQ). The board discussed volume-related policy issues at their October 2000 board meeting with three participants of the IOM workshop, Colin Begg, R. Adams Dudley, and Edward Hannan. This White Paper briefly summarizes the findings from the workshop and presents board recommendations for implementing volume-based policies relevant to cancer care.

THE VOLUME–OUTCOME RELATIONSHIP IN THE CONTEXT OF HEALTH CARE QUALITY MEASUREMENT

The National Cancer Policy Board concluded in its 1999 report *Ensuring Quality Cancer Care* that, based on the best available evidence, some individuals with cancer do not receive care known to be effective for their conditions. The magnitude of the problem is not known, but the board believes it is substantial. Evidence points to underuse of some interventions known to be effective (e.g., radiation therapy, adjuvant chemotherapy after surgery), overuse of interventions for which evidence supports alternative interventions (e.g., mastectomy versus breast conserving surgery), and misuse of effective interventions (e.g., administering inappropriate doses of chemotherapy). Despite compelling evidence of quality problems, it is difficult for individual consumers, health care purchasers, and others to make informed choices about cancer care, in part because the data needed to provide quality information specific to a particular physician or hospital are generally not available (IOM, 2000b). To ascertain

whether practitioners are providing appropriate radiation and adjuvant chemotherapy after surgery, for example, one would have to assemble data from hospitals, outpatient settings, and possibly patients themselves (e.g., to ascertain treatment preferences). Such data may be examined as part of a health services research project or within a specific care system, but they are generally not available regionally or nationally. In the absence of good data on processes of care, data about outcomes (e.g., mortality, functional status) 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. Such risk-adjusted outcomes data are, however, not generally available to assess the quality of cancer care because of the time and expense associated with gathering and interpreting clinical data.

When data on processes and outcomes of care are not available, alternative indicators may be used to ascertain quality. Health services researchers have assessed whether the site at which care is delivered is predictive of outcomes by examining associations between aspects of the organization and delivery of cancer care and health outcomes. Available evidence is insufficient to say that cancer care is better or worse when offered by specialized compared to generalist facilities or providers, or in managed care versus fee-for-service environments (IOM, 1999). Evidence is compelling, however, for a strong positive association between the volume of certain types of cancer care and better outcomes. Assessments of the volume–outcome relationship have tended to focus on surgical interventions because hospital data are generally available to study surgical procedures and their associated short-term mortality. It is more difficult to study the relationship between volume and outcomes for other types of interventions (e.g., chemotherapy, medical management) because there are insufficient sources of data on care administered outside of hospitals and on specific processes of care (IOM, 2000b). Likewise, there are virtually no widely available sources of information on longer-term outcomes of care such as quality of life and functional status. What follows is a summary of the literature on cancer surgery and the volume–outcome relationship.

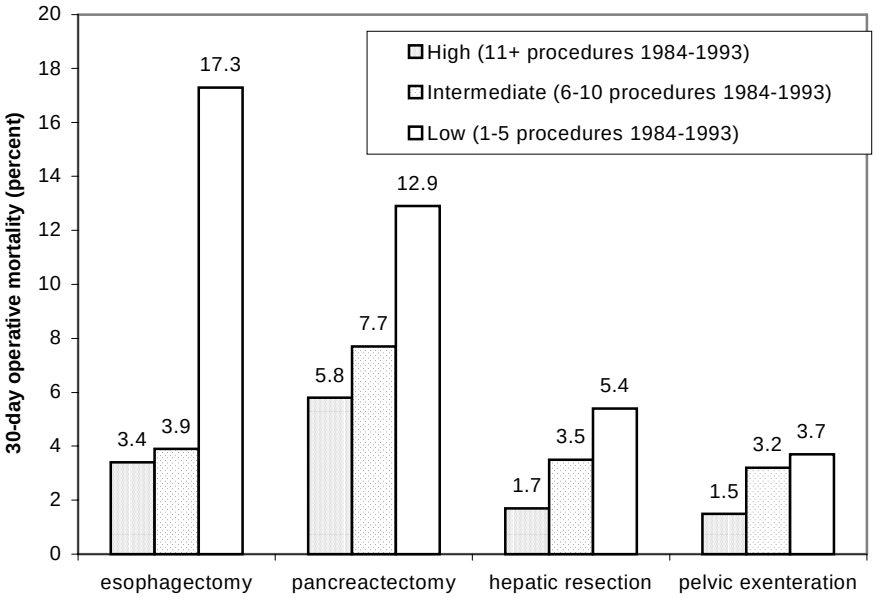
EVIDENCE OF A VOLUME–OUTCOME RELATIONSHIP FOR CANCER INTERVENTIONS

As part of their synthesis of the literature on the volume–outcome relationship, Dr. Halm and colleagues reviewed 20 population-based studies of surgical interventions for cancer (the section of the literature review pertaining to cancer is included in Appendix A). The studies varied in their definition of high and low volume, whether hospital or physician volume (or both) were assessed, and how differences in characteristics of patients in high- and low-volume hospitals were taken into consideration in the analysis (i.e., case-mix adjustment). Despite differences in study design and methods, there is consistency in the published

results—a higher-volume–better-outcome association was observed in all but three of the studies reviewed (these three studies showed no volume–outcome association). Almost all studies focused on short-term postoperative mortality (i.e., either in-hospital or 30-day mortality) and surgical complications, but a few investigators extended the time to follow-up. Birkmeyer and colleagues, for example, found a fourfold increase in in-hospital mortality rates following pancreaticoduodenectomy performed for Medicare patients in low- as compared to high-volume hospitals (16 versus 4 percent) (Birkmeyer et al., 1999a). A follow-up study of these patients showed that the significant volume-related mortality advantage persisted at 3 years post-surgery (37 versus 25 percent in low- versus high-volume hospitals) (Birkmeyer et al., 1999b).

The volume–outcome relationship appears to be particularly strong 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 generally at least two to three times greater in low- versus high-volume hospitals. Operative mortality rates by volume for four high-risk, cancer-related surgical procedures performed among Medicare beneficiaries are shown in Figure 1 (Begg et al., 1998). Although there is a statistically significant trend confirming improved outcomes with higher volume, providers in the intermediate-volume group are sometimes indistinguishable from either low- or high-volume providers.

FIGURE 1 Impact of hospital volume on operative mortality for major cancer surgery among Medicare beneficiaries.
SOURCE: Begg et al., 1998.



For other procedures or conditions under review, the volume effect was not as great or as consistent. For common cancer-related surgical procedures (e.g., surgery for colorectal cancer), some studies show no effect, whereas others show statistically significant, but relatively small, effects. In a recent study by Hannan, for example, adjusted mortality rates were from 2 to 7 percentage points higher for low- compared to high-volume hospitals performing colectomy, lobectomy, and gastrectomy (Hannan, in press). In the few studies in which the effects of both surgeon and hospital volume have been assessed, only hospital volume is consistently related to better outcomes.

INTERPRETING THE VOLUME–OUTCOME RELATIONSHIP

Volume is recognized as an imperfect correlate of quality. Volume per se does not result in good outcomes in health care but is instead a proxy measure for other factors that affect care. These factors might include physician skill, experienced interdisciplinary teams, or well-organized care processes. However, with few exceptions, the literature does not shed light on the structures or processes of care that underlie the apparent relationship.

A strength of volume as an indicator of quality of care is that it is relatively easy to obtain from available administrative databases. For procedures performed infrequently (e.g., esophagectomy), it is very difficult to measure quality directly (i.e., using physician- or hospital-specific outcomes data) because of the instability of small numbers (e.g., a few deaths can greatly influence annual rates); thus proxies such as volume may have to suffice. Here, one could base selective referral on minimum volume standards. For conditions in which higher volume has been shown to improve outcomes and for which caseloads are large enough to support outcomes measurement, it is also feasible to assess quality based on both outcomes data and volume. The use of a combination of data sources is preferable to using any one source alone. Assessing multiple indicators of quality is also preferable to relying on any one indicator alone (IOM, 2000a).

Coupling volume data with clinical data could lead to the identification of processes and structures of care that distinguish high- and low-volume providers and that predict outcomes. The relative contributions of volume, process, and structure of care can be assessed only when comprehensive data are collected systematically as part of a special registry. In New York and New Jersey (and soon in California), for example, statewide clinical databases are available for cardiac surgery that allow analyses of outcomes by both individual surgeons and hospitals. Cancer registries are available for surveillance purposes, but they usually lack sufficient clinical information for quality-of-care studies (IOM, 2000b).

Volume, when used as an indicator of quality, can be imprecise. Even though, in the aggregate, high- compared to low-volume providers have better outcomes, there is some variation so that not all high-volume providers have

better outcomes and not all low-volume providers have worse outcomes. Consequently, the quality of care offered by any particular provider cannot be predicted accurately with information on volume alone. Furthermore, most volume studies to date have focused on short-term outcomes and on mortality. Whether outcomes such as quality of life or functional status improve with higher volume is not known.

Despite its apparent value as a quality indicator, especially for low-frequency care, there are a number of unresolved issues that make volume difficult to operationalize in the context of health care quality improvement programs:

- When volume effects have been noted, it is unclear where along the volume continuum a threshold exists, above which outcomes are better but do not continue to improve with further volume increases.
- Studies generally do not illuminate how experience with procedures that are closely related to the procedure under study affect outcomes (e.g., should esophagectomies performed for indications other than cancer "count" toward volume?).
- It is likely that effects of physician and hospital volume combine or interact. The relative contributions of physician and hospital volume to outcomes, however, have been examined in only a few studies.
- Once high volume is attained, does it have to be sustained, or can lower volumes be adequate to maintain good performance?

POTENTIAL IMPACT OF POLICIES TO CONCENTRATE CANCER CARE IN HIGH-VOLUME HOSPITALS

There are roughly 5,000 community hospitals in the United States and virtually all of them provide at least some cancer care (AHA, 2000). Cancer-related surgeries for which the relation between volume and outcome appears to be strongest are performed infrequently: in 1997, there were an estimated 2,011 cancer-related esophagectomies and 3,832 pancreatectomies performed in the United States (Table 1). Relatively few hospitals would likely be affected by policies involving these uncommon procedures because no more than one-quarter of hospitals perform such surgeries. If a volume–outcome effect were established for more common procedures (e.g., gastrectomy, lobectomy, colectomy), a larger share of hospitals (up to 80 percent) would likely be affected by volume-based policies, such as selective referral programs. Selective referral programs might be difficult to implement for infrequently performed procedures because of the limited number of hospitals that have high volumes of these procedures: in 1997, an estimated 37 hospitals nationwide performed seven or more esophagectomies and 85 hospitals had this volume of pancreatectomies. The

TABLE 1 Distribution of Selected Cancer-Related Surgical Procedures by Hospital Volume, United States, 1997

Procedure[a]	Number of Discharges		Hospitals Performing at Least One Procedure				Definition of Lower Volume[b]	Percentage of Discharges from Lower-Volume Hospitals	
			Number		*Percentage*				
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Number	Estimate	95% CI
Total			5,113		100.0				
Esophagectomy	2,011	1,626–2,396	869	772–966	17.0	15.1–18.9	< 3	41.9	32.3–51.4
Pancreatectomy	3,832	2,808–4,857	1,252	1,141–1,363	24.5	22.3–26.7	< 3	30.2	21.1–39.3
Gastrectomy	10,592	9,744–11,440	2,526	2,410–2,642	49.4	47.1–51.7	< 5	34.5	30.1–38.9
Lobectomy	27,763	25,056–30,470	2,443	2,334–2,552	47.8	45.6–49.9	< 10	20.1	17.0–23.2
Colectomy	86,676	82,518–90,834	4,165	4,053–4,277	81.5	79.3–83.6	< 20	23.2	21.0–25.5

NOTE: CI = confidence interval; ICD-9-CM = *International Classification of Diseases, Ninth Edition*, Clinical Modifications.

[a]Esophagectomy procedures defined as ICD-9-CM diagnostic codes = 150.x; procedure codes = 42.4x, 42.5x, 42.6x. Pancreatectomy defined as ICD-9-CM diagnostic codes = 157.x; procedure codes = 52.51, 52.53, 52.59, 52.6, 52.7. Gastrectomy defined as ICD-9-CM diagnostic codes = 151.x; procedure codes = 43.5–43.99. Lobectomy of lung defined as ICD-9-CM diagnostic codes = 162.2–162.9; procedure codes = 32.4. Colectomy defined as ICD-9-CM diagnostic codes = 153.x; procedure codes = 45.73, 45.74, 45.75, 45.76.

[b]There are no established definitions of "low" volume. These cutpoints are presented for illustration only.

SOURCE: Special tabulations: NCPB staff; Healthcare Cost and Utilization Project (HCUP), 1997 National Impatient Sample, Release 6, Agency for Healthcare Research and Quality, 1999.

potential impact of volume-based policies on outcomes for these two procedures appears to be substantial because a relatively large share of discharges are from very low-volume hospitals (i.e., in 1997, an estimated 42 percent of esophagectomies and 30 percent of pancreatectomies were performed in hospitals with fewer than three procedures per year [Table 1]). Relatively few cancer-related esophagectomies and pancreatectomies (5 and 4 percent of discharges for these procedures, respectively) are performed in hospitals located in nonmetropolitan areas according to the 1997 Healthcare Cost and Utilization Project analyses.

In a study of the potential impact in California of selective referral to high-volume hospitals, Dudley and colleagues (2000) estimated that in 1997, 27 deaths associated with esophagectomy and pancreatectomy performed in low-volume hospitals could have been averted with care in a high-volume hospital.

NATIONAL CANCER POLICY BOARD RECOMMENDATIONS

It is often difficult to judge when to implement policies based on research findings. The board considered four criteria to assess the strength of the evidence on the volume–outcome relationship and its adoption as a criterion for referral:

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

The board concluded that these criteria are met for two procedures included in the literature review—surgery for cancer of the pancreas and esophagus. The board chose to limit its recommendation to these two surgical procedures because of the size of the relationship and the consistency of the findings in the literature. Although the board found evidence regarding other procedures compelling (e.g., removal of pelvic organs, complex chemotherapy), it concluded that initial applications of cancer-related volume measures in quality assurance and improvement programs should be limited to those areas in which the body of evidence is robust. Furthermore, the board concluded that when research confirms a volume–outcome link, information should be disseminated to the public to support health care decision-making. In making such a disclosure, however, the limitations of the data and how to interpret them must be clear for their intended audience.

The board proposes the following two recommendations to incorporate well-validated volume measures into quality assurance and improvement programs and to support further research on the volume–outcome relationship and its value in improving the quality of cancer care.

Recommendation 1: When a large and significant volume–outcome relationship is established firmly by the literature through consistent findings in multiple studies (i.e., esophagectomy, pancreatectomy), volume should be incorporated as a quality indicator into ongoing quality-of-care programs and initiatives. Examples of such applications include the following:

1. Public and private health care purchasers' use of quality indicators (e.g., selective referral programs, consumer education);

2. Health plans' and providers' internal quality assurance monitoring;

3. Quality assurance organizations' surveillance activities:

- **The Health Care Financing Administration's Peer Review Organizations' (PROs') assessment of the quality of care for Medicare beneficiaries**
- **Joint Commission on Accreditation of Healthcare Organizations' surveys of hospital and other health care organizations**
- **National Committee for Quality Assurance reporting of quality indicators for managed care health plans**

4. Professional societies' assessments of patterns of care (e.g., American College of Surgeons' Commission on Cancer [ACS-CoC])

5. Consumer groups' campaigns to educate the public on quality-of-care issues.

Well-validated quality measures can be applied in a variety of settings. Health insurance purchasers could use findings from research on the volume–outcome relationship to stipulate "evidence-based referrals" in contracts with health plans. The Pacific Business Group on Health (PBGH), a large purchasing coalition, is negotiating with health plans with which it contracts to increase the proportion of patients with selected conditions who are treated at high-volume hospitals (e.g., individuals with esophageal cancer in need of esophagectomies are to be referred to hospitals performing at least seven such procedures each year). PBGH also provides condition-specific volume data for all California hospitals on its consumer website (www.healthscope.org) along with guidance on how to interpret the data.

Employers could also make information about the relationship between volume and outcome available to employees directly (e.g., on a company intranet site) and could encourage employees to choose hospitals and providers based on available evidence. Similarly, health plans could direct members to high-volume providers. 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).

Information about volume can be applied without public disclosure, for example, within systems of care for quality improvement programs. HCFA, for example, could use volume data to target interventions of its PROs that operate at the state level to ensure the quality of care for Medicare beneficiaries. Low-volume providers may withdraw voluntarily to avoid scrutiny, or they may be motivated to achieve minimal volume standards.

Recommendation 2: Federal and private research sponsors such as the National Cancer Institute, the Agency for Healthcare Research and Quality, health care purchasers, health plans, and provider groups, through public–private partnerships, should support program evaluation and research projects to: (1) elucidate the nature of the volume–outcome relationship and its application to quality improvement, and (2) monitor the implementation (and effects) of volume-based policies.

Much remains to be known about the relationship between volume and outcomes in the context of cancer care. Although a number of databases exist with which to assess the relationship, they have not been used extensively to assess cancer care, nor have evaluations been planned of ongoing efforts to integrate volume-based measures into quality improvement programs. A wide-ranging research agenda—from policy research to basic methodological research—is necessary to better understand the relationship between volume and outcome and how best to implement policies to improve care.

Elucidating the Nature of the Volume–Outcome Relationship and Its Application to Quality Improvement

Research is needed to determine the range of cancer care for which a volume–outcome relationship exists. This could be accomplished through a systematic and comprehensive examination of the relationship for both surgical and nonsurgical interventions, using existing data resources (e.g., AHRQ's HCUP database, state hospital discharge files, cancer registries, ACS-CoC's and the American Cancer Society's National Cancer Data Base). Such research will help determine the need for condition- or procedure-specific, prospective, population-based clinical databases and registries. Clinical databases and registries may be needed for more common cancer-related interventions to examine factors that mediate volume–outcome relationships.

Monitoring the Implementation (and Effects) of Volume-Based Policies

Several concerns have been raised regarding the adoption of volume-based quality measures, so methods are needed to monitor the impact of the adoption of such measures. A major concern is that selective referral programs might run counter to patient preferences for care close to home. Access to high-volume providers might be especially difficult for residents of rural areas and for those who lack resources to travel to hospitals that are far away. For some procedures, it may never be possible to regionalize care fully because some patients may need immediate treatment or be too unstable to transfer to a higher-volume setting. Where low-volume services have been closed, patients may experience a loss of access to a range of services, not just the procedure for which a volume–outcome relationship is known. Furthermore, providers may lose the ability to appropriately manage the postsurgical complications that arise in patients who have been referred to a distant high-volume hospital, but who return home for follow-up care.

There are also potential effects on area marketplace structure and competition, such as the increased market power of high-volume hospitals (e.g., prices could rise), or barriers to the entry of new competitors (i.e., it is difficult to start at high volume). Furthermore, there is a potential for unintended consequences of a selective referral program—there could be medically inappropriate admissions to boost volumes to meet cutoffs. Also unexplored is a potential decrement in quality at very high volumes as a consequence of selective referral programs.

Mechanisms are needed to monitor these and other effects of volume-based policies. One resource to assess the impact of such policies is the AHRQ's Healthcare Cost and Utilization Project (HCUP). The HCUP National Inpatient Sample in 1997 included information on 7.1 million discharges from a 20 percent sample of U.S. community hospitals (1,012 hospitals in 22 states). Because a hospital's total discharges are available, the annual volume of any particular procedure can be tallied by diagnosis.

Preliminary analyses of HCUP data suggest that volume-based policies for low-frequency procedures such as esophagectomy and pancreatectomy might involve relatively few hospitals, but could have great impact on outcomes because many of these procedures appear to be performed in very-low-volume hospitals. Surveillance data from HCUP (and other sources) should be scrutinized by organizations that could use the information to implement programs to target interventions to areas where care remained concentrated in lower-volume settings. Referral patterns for medical care are very difficult to change, and the efforts of a number of groups will be required to foster the concentration of selected care in higher-volume settings. HCUP includes states' hospital discharge data, a valuable data source for health services research, but the database has certain limitations (e.g., lacks detailed clinical data, is not longitudinal, provides

no information on patient preferences). Additional sources of data will be needed to fully evaluate the impact of volume-based policies.

Research is also necessary to assess consumer and provider response to volume-based quality indicators—their interest in, and interpretation of, volume as an indicator of health care quality relative to other measures.

Technical issues will have to be resolved through implementation-related research. Operationalizing volume-based quality indicators will require agreement on the definition of the conditions and procedures to be included in the measure (e.g., ICD-9 codes); volume thresholds or cutpoints to identify high- and low-volume hospitals, standard methods to measure and monitor hospital volume (e.g., annual or biannual measurement, effects of hospital mergers and affiliations on categorization), and appropriate reporting formats for health care consumers.

REFERENCES

American Hospital Association (AHA), "Fast Facts on U.S. Hospitals from Hospital Statistics," www.aha.org/resource/newpage.html, accessed September 12, 2000.

Begg CB, Cramer LD, Hoskins WJ, Brennan MF, 1998 "Impact of Hospital Volume on Operative Mortality for Major Cancer Surgery," *JAMA* 280(20): 1747–1751.

Birkmeyer JD, Finlayson SR, Tosteson AN, et al., 1999a, "Effect of Hospital Volume on In-Hospital Mortality with Pancreaticoduodenectomy," *Surgery* 125(3): 250–256.

Birkmeyer JD, Warshaw AL, Finlayson SR, et al., 1999b, "Relationship Between Hospital Volume and Late Survival After Pancreaticoduodenectomy," *Surgery* 126(2):178–183.

Dudley RA, Johansen KL, Brand R, et al., 2000, "Selective Referral to High-Volume Hospitals: Estimating Potentially Avoidable Deaths," *JAMA* 283(9):1159–1191.

Hannan EL, Radzyner M, Rubin D, et al. (*Surgery*, in press), "The Influence of Hospital and Surgeon Volume on Mortality for Common Cancer Procedures."

IOM, Hewitt M, Simone JV (editors), 1999, *Ensuring Quality Cancer Care*, National Academy Press, Washington DC.

IOM, Hewitt M, 2000a, *Interpreting the Volume–Outcome Relationship in the Context of Health Care Quality: Workshop Summary*, National Academy Press, Washington DC.

IOM, Hewitt M, Simone JV (editors), 2000b, *Enhancing Data Systems to Improve the Quality of Cancer Care*, National Academy Press, Washington DC.

Appendix A
Volume and Outcome in Cancer Surgery

Excerpted from "How Is Volume Related to Quality in Health Care? A Systematic Review of the Research Literature," by Ethan A. Halm, MD, MPH, Clara Lee, MD, MPP, and Mark R. Chassin, MD, MPP, MPH, Department of Health Policy, Mount Sinai School of Medicine

Prepared for:
Institute of Medicine, National Academy of Sciences
Division of Health Care Services,
Committee on Quality of Care in America
National Cancer Policy Board

Workshop
Interpreting the Volume–Outcome Relationship
in the Context of Health Care Quality
May 1, 2000

VOLUME AND OUTCOME IN CANCER SURGERY

We examined a total of 38 studies on cancer (criteria for rating the quality of published studies and literature review methods are described on pages 22–25). 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 (see attached summaries). The three articles that looked at other cancer procedures are summarized separately in a table 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, 1998). 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 (Gordon, 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 (≤ 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% versus 19%. When surgeon volume and hospital volume were analyzed together, however, only hospital volume was significant.

Sosa et al. (1998) 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 esophagectomy 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 ($\geq$37 cases per year) was 1.65% higher than at high-volume hospitals ($\geq$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 United States. 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 one study (Gordon, 1999), the physician in one study (Porter, 1998), and both hospitals and physicians in three (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 recurrence (hazard ratio = 1.80). High-volume surgeons were more likely to perform a low ante-

rior 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 20 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 methodological 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 underlie 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.

CRITERIA FOR RATING THE QUALITY OF PUBLISHED STUDIES

We developed a scoring system to assess the quality of the research studies included in our systematic review. The full list of criteria is described on page 23. 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 $\geq$ 50 for scoring this criterion. If the total sample size was 1,000 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 literature 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 1 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 1 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 1 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 1 point. If 2 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

We 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 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 conditions with the term regionalization.

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 paper includes the findings of our review of cancer-related procedures and conditions. Study 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. Our final criteria for quality assessment and the scoring system were described earlier and are listed in on page 23. 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 two reviewers.

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/or MD < 50		H ≥ 20 and MD ≥50		0	1	2	
3. Total sample size (cases)	< 1000		≥ 1000		0	1		
4. Number 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 separately		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

PANCREAS

Study	Population	Time period	Patient #	MD #	Hospital #	Unit of analysis	Primary outcome	Risk adjust-ment 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	3yr death	Admin	Hosp Very low: < 1 Low: 1–2 High: ≥ 5	OR = 0.69	7
Birkmeyer 1999b	Medicare US Benign & malignant	1992–1995	7229	NS	1772	Hosp	Inpt death 30d death	Admin	Hosp Very low: < 1 Low: 1–2 High: ≥ 5	Inpt death: 16% vs. 4.1% (very high 1.7%) 30d death: 12.9 vs. 3.0%	7
Sosa 1998a	All Maryland	1990–1995	1236	373	48	MD Hosp Both	Inpt death	Admin	MD: Low: < 5, High: > 50 Hosp: Low: < 5 High: ≥ 20	LVH vs. HVH: RR = 19.3 HVH better, regardless of MD volume	9

(continued)

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 64d 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	1991–1994	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: 19vs. 2.2% (RR = 8.7)	6

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

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

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

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 in press	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- rela-tive to low-volume): OR = 0.6 Pneumonec-tomy: 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

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 in press	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 vs. LVS; OR = .64; Hosp: HVH vs. 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 cancer	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 LVH: low-volume hospital HVH: high-volume hospital DSS: disease-specific survival HR: hazards ratio
LVP: low-volume physician HVP: high-volume physician MRP: medium-volume physician LAR: low anterior resection NS: not specified

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 in press	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 HVP: high-volume physician NS: not specified
LVH: low-volume hospital HVH: high-volume hospital

REFERENCES

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.

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.

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.

Glasgow RE, Mulvihill SJ. Hospital volume influences outcome in patients undergoing pancreatic resection for cancer. *Western Journal of Medicine*. 1996;165(5):294–300.

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.

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.

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.

Hannan E. The influence of hospital and surgeon volume on mortality for common cancer procedures. *Surgery*, in press.

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.

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.

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.

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.

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.

Porter GA, Soskolne CL, Yakimets WW, Newman SC. Surgeon-related factors and outcome in rectal cancer. *Ann Surg*. 1998;227(2):157–67.

Romano PS, Mark DH. Patient and hospital characteristics related to in-hospital mortality after lung cancer resection. *Chest*. 1992;101(5):1332–7.

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.

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.

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.

Sosa JA, Bowman HM, Gordon TA, et al. Importance of hospital volume in the overall management of pancreatic cancer. *Annals of Surgery*. 1998;228(3):429–38.

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.