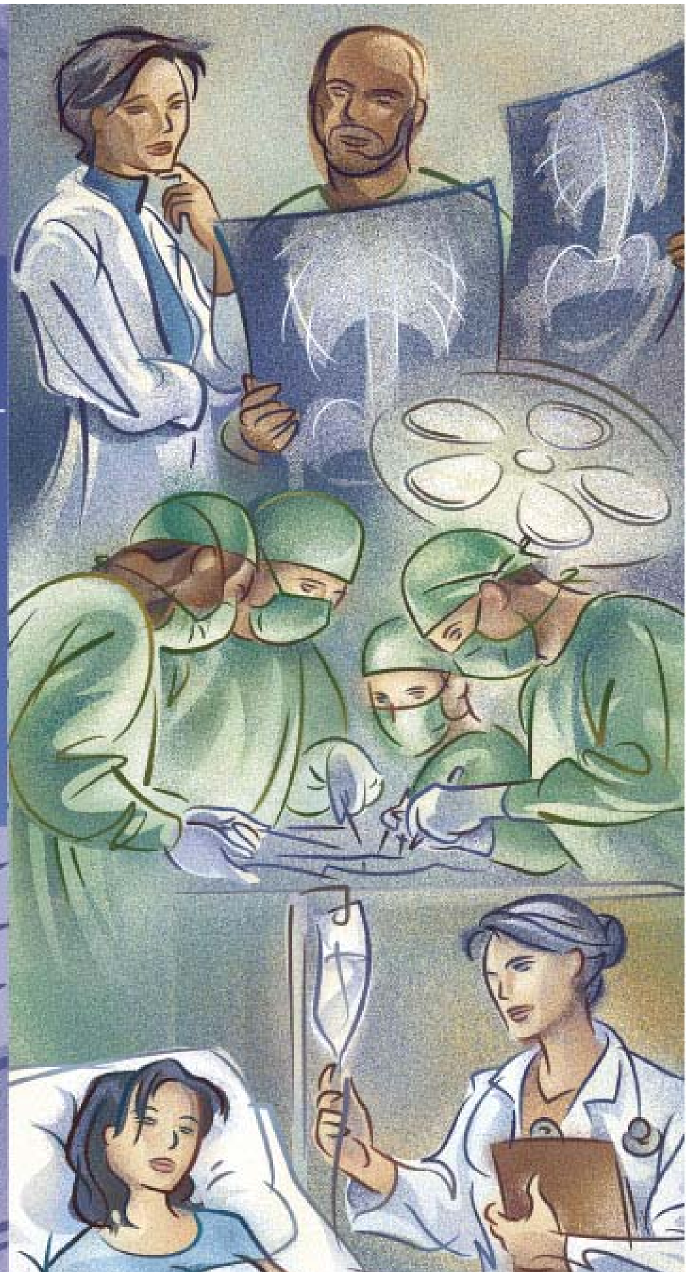
AHRQ QUALITY INDICATORS

Guide to Inpatient Quality Indicators



AHRQ Quality Indicators

Guide to Inpatient Quality Indicators:
Quality of Care in Hospitals – Volume, Mortality, and Utilization

Department of Health and Human Services
Agency for Healthcare Research and Quality
http://www.qualityindicators.ahrq.gov

June 2002
Version 3.0 (February 20, 2006)

Preface

In health care as in other arenas, that which cannot be measured is difficult to improve. Providers, consumers, policy makers, and others seeking to improve the quality of health care need accessible, reliable indicators of quality that they can use to flag potential problems or successes; follow trends over time; and identify disparities across regions, communities, and providers. As noted in a 2001 Institute of Medicine study, *Envisioning the National Health Care Quality Report*, it is important that such measures cover not just acute care but multiple dimensions of care: staying healthy, getting better, living with illness or disability, and coping with the end of life.

The Agency for Healthcare Research and Quality (AHRQ) Quality Indicators (QIs) are one Agency response to this need for multidimensional, accessible quality indicators. They include a family of measures that providers, policy makers, and researchers can use with inpatient data to identify apparent variations in the quality of inpatient or outpatient care. AHRQ's Evidence-Based Practice Center (EPC) at the University of California San Francisco (UCSF) and Stanford University adapted, expanded, and refined these indicators based on the original Healthcare Cost and Utilization Project (HCUP) Quality Indicators developed in the early 1990s.

The AHRQ QIs are organized into four modules: **Prevention Quality Indicators (PQIs)**, **Inpatient Quality Indicators (IQIs)**, **Patient Safety Indicators,** and **Pediatric Quality Indicators (PDIs).** AHRQ has published the modules as a series. The first module—Prevention Quality Indicators—was released in 2001 and is available at AHRQ's Quality Indicators Web site at http://www.qualityindicators.ahrq.gov.

This second module focuses on health care provided within the inpatient hospital setting. The Inpatient Quality Indicators include three distinct types of measures. **Volume** measures examine the volume of inpatient procedures for which a link has been demonstrated between the number of procedures performed and outcomes such as mortality. **In-hospital mortality** measures examine outcomes following procedures and for common medical conditions. **Utilization** examines procedures for which questions have been raised about overuse, underuse, and misuse.

Full technical information on the first two modules can be found in *Evidence Report for Refinement of the HCUP Quality Indicators*, prepared by the UCSF-Stanford EPC. It can be accessed at AHRQ's Quality Indicator Web site (http://www.qualityindicators.ahrq.gov). The third module—Patient Safety Indicators (PSIs)—was released in May 2003. Information on the PSIs, including the technical information, software and other documentation is also available at AHRQ's Quality Indicators Web site.

Improving the quality of inpatient hospital services is a critical part of efforts to provide high quality health care in the United States. This guide is intended to facilitate such efforts. As always, we would appreciate hearing from those who use our measures and tools so that we can identify how they are used, how they can be refined, and how we can measure and improve the quality of the tools themselves. You may contact us by sending an e-mail to support@qualityindicators.ahrq.gov.

Irene Fraser, Ph.D., Director
Center for Organization and Delivery Studies

The programs for the Inpatient Quality Indicators (IQIs) can be downloaded from http://www.qualityindicators.ahrq.gov/iqi_download.htm. Instructions on how to use the programs to calculate the IQI rates are contained in the companion text, *Inpatient Quality Indicators: Software Documentation (both SAS and SPSS).*

Acknowledgments

Support efforts, including refinement and enhancement of the AHRQ Quality Indicators and related products, are provided by the Support for Quality Indicators-II contract team.

The following individuals from **Battelle Memorial Institute, Stanford University,** and **University of California (UC)** constitute the Support for Quality Indicators-II core team:

Sheryl M. Davies, M.A.
Bruce Ellis, M.S.
Jeffrey Geppert, J.D.
Mark Gritz, Ph.D.
Greg Hubert, B.S.
Elaine Keller, M.Ed.
Kathryn M. McDonald, M.M.
Patrick Romano, M.D., M.P.H
Jeff Schoenborn, B.S.

The **Agency for Healthcare Research and Quality** Support for Quality Indicators team includes:

Marybeth Farquhar, Project Officer
Mamatha Pancholi, Project Officer
Mary B. Haines, Contract Officer

The following staff from the **Evidence-based Practice Center (EPC) at UCSF-Stanford** performed the evidence review, completed the empirical evaluation, and created the programming code and technical documentation for the AHRQ Inpatient Quality Indicators:

Core Project Team

Mark McClellan, M.D., Ph.D., *principal investigator*
Kathryn M. McDonald, M.M., *EPC coordinator*
Sheryl M. Davies, M.A.
Jeffrey Geppert, J.D.
Patrick Romano, M.D., M.P.H.
Kaveh G. Shojania, M.D.

Other Contributors

Amber Barnato, M.D.
Paul Collins, B.A.
Bradford Duncan M.D.
Michael Gould, M.D., M.S.
Paul Heidenreich, M.D.
Corinna Haberland, M.D.
Paul Matz, M.D.
Courtney Maclean, B.A.
Susana Martins, M.D.
Kristine McCoy, M.P.H.
Suzanne Olson, M.A.
L. LaShawndra Pace, B.A.
Mark Schleinitz, M.D.
Herb Szeto, M.D.
Carol Vorhaus, M.B.A
Peter Weiss, M.D.
Meghan Wheat, B.A.

Consultants

Douglas Staiger, Ph.D.

The following staff from **Social & Scientific Systems**, **Inc.** developed the original software product, documentation, and guide:

Programmers

Leif Karell
Kathy McMillan

Technical Writer

Patricia Burgess

Graphics Designer

Laura Spofford

Contributors from the **Agency for Healthcare Research and Quality**:

Anne Elixhauser, Ph.D.
Denise Remus, Ph.D., R.N.
H. Joanna Jiang, Ph.D.
Margaret Coopey, R.N., M.G.A, M.P.S.

The contribution of the peer reviewers of the evidence report and the beta-testers of the software products is also acknowledged, their input was invaluable.

Table of Contents

List of Tables

1.0 Introduction to the AHRQ Inpatient Quality Indicators

Hospitals in the United States provide the setting for some of life's most pivotal events—the birth of a child, major surgery, treatment for otherwise fatal illnesses. These hospitals house the most sophisticated medical technology in the world and provide state-of-the-art diagnostic and therapeutic services. But access to these services comes with certain costs. About 30% of personal health care expenditures in the United States go towards hospital care,[1] and the rate of growth in spending for hospital services has only recently leveled out after several years of increases following a half a decade of declining growth.[2] Simultaneously, concerns about the quality of health care services have reached a crescendo with the Institute of Medicine's series of reports describing the problem of medical errors[3] and the need for a complete restructuring of the health care system to improve the quality of care.[4] Policymakers, employers, and consumers have made the quality of care in U.S. hospitals a top priority and have voiced the need to assess, monitor, track, and improve the quality of inpatient care.

Hospital administrative data offer a window into the medical care delivered in our nation's hospitals. These data, which are collected as a routine step in the delivery of hospital services, provide information on diagnoses, procedures, age, gender, admission source, and discharge status. From these data elements, it is possible to construct a picture of the quality of medical care. Although quality assessments based on administrative data cannot be definitive, they can be used to flag potential quality problems and success stories, which can then be further investigated and studied. Hospital associations, individual hospitals, purchasers, regulators, and policymakers at the local, State, and Federal levels can use readily available hospital administrative data to begin the assessment of quality of care. The AHRQ Quality Indicators (QIs) are a tool that takes advantage of hospital administrative data. The Inpatient Quality Indicators (IQIs) represent the current state-of-the-art in measuring the quality of hospital care through analysis of inpatient discharge data.

The AHRQ QIs are now being used for applications beyond quality improvement. In 2003, AHRQ first published the *National Healthcare Quality Report*[5] (NHQR) and *National Healthcare Disparities Report*[6] (NHDR) which provide a comprehensive picture of the level and variation of quality within four components of health care quality—effectiveness, safety, timeliness, and patient centeredness. These reports incorporated many Prevention Quality Indicators, Inpatient Quality Indicators, and Patient Safety Indicators. Selected mortality and utilization indicators from the IQI module will be included in the next NHQR and NHDR reports.[7] Some organizations have used the AHRQ Quality Indicators to produce web based, comparative reports on hospital quality, such as the Texas Department of State Health Services[8] and the Niagara Coalition[9]. These organizations also supplied users with guidance on indicator interpretation. Other organizations have incorporated selected AHRQ QIs into pay for performance demonstration projects or similar programs, such as the Centers for Medicare and Medicaid Services (CMS)[10] and Anthem Blue Cross Blue Shield of Virginia where hospitals would be financially rewarded for

---

[1]. http://www.cms.hhs.gov/NationalHealthExpendData/downloads/nheprojections2004-2014.pdf: Table 2 National Health Expenditure Amounts, and Annual Percent Change by Type of Expenditure: Selected Calendar Years 1998-2014.
[2]Strunk BC, Ginsburg PB, Gabel JR. Tracking Health Care Costs. Health Affairs, 26 September 2001 (Web exclusive).
[3]Institute of Medicine. To Err is Human: Building a Safer Health System. Kohn LT, Corrigan JM, Donaldson MS (eds.) Washington DC: National Academy Press, 2000.
[4]Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Committee of Quality of Care in America. Washington DC: National Academy Press, 2001.
[5] Agency for Healthcare Research and Quality. *National Healthcare Quality Report*. Rockville, MD, U.S. Department of Health and Human Services, Agency for Healthcare Research and Quality, December 2003.
[6] Agency for Healthcare Research and Quality. *National Healthcare Disparities Report*. Rockville, MD, U.S. Department of Health and Human Services, Agency for Healthcare Research and Quality, July 2003.
[7] The 2005 NHQR and NHDR reports are available at http://www.qualitytools.ahrq.gov/.
[8] Texas Center for Health Statistics. *Indicators of Inpatient Care in Texas Hospitals, 2003*. http://www.dshs.state.tx.us/THCIC/Publications/Hospitals/IQIReport2003/IQIReport2003.shtm. Accessed January 2006.
[9] Niagara Health Quality Coalition. *2005 New York State Hospital Report Card.* http://www.myhealthfinder.com/newyork05/glancechoose.htm Accessed January 2006.
[10] Centers for Medicare & Medicaid Services. *The Premier Hospital Quality Incentive Demonstration*. http://www.cms.hhs.gov/HospitalQualityInits/downloads/HospitalPremierFactSheet.pdf. Accessed January 2006.

performance. Guidance on these alternative uses of the AHRQ QIs is summarized in an AHRQ *Summary Statement on Comparative Reporting*[11] and accompanying publication titled G*uidance for Using the AHRQ Quality Indicators for Hospital-Level Public Reporting or Payment*[12].

This update of the AHRQ IQIs (Version 3.0), reflects changes in indicators associated with ICD-9-CM coding updates for FY 2006 (effective 10-1-2005).

1.1 What Are the Inpatient Quality Indicators?

The IQIs are a set of measures that can be used with hospital inpatient discharge data to provide a perspective on quality and include the following:

- **Volume** indicators are proxy, or indirect, measures of quality. They are based on evidence suggesting that hospitals performing more of certain intensive, high-technology, or highly complex procedures may have better outcomes for those procedures. Volume indicators simply represent counts of admissions in which these procedures were performed.

- **Mortality indicators for inpatient procedures** include procedures for which mortality has been shown to vary across institutions and for which there is evidence that high mortality may be associated with poorer quality of care.

- **Mortality indicators for inpatient conditions** include conditions for which mortality has been shown to vary substantially across institutions and for which evidence suggests that high mortality may be associated with deficiencies in the quality of care.

- **Utilization** indicators examine procedures whose use varies significantly across hospitals and for which questions have been raised about overuse, underuse, or misuse. High or low rates for these indicators are likely to represent inappropriate or inefficient delivery of care.

[11] AHRQ Summary Statement on Hospital Public Reporting. http://www.qualityindicators.ahrq.gov/news/AHRQSummaryStatement.pdf,

[12] Remus D, Fraser I. *Guidance for Using the AHRQ Quality Indicators for Hospital-level Public Reporting or Payment*. Rockville, MD: Department of Health and Human Services, Agency for Healthcare Research and Quality; 2004. AHRQ Pub. No. 04-0086-EF. The document may be downloaded from the AHRQ Quality Indicator website at http://www.qualityindicators.ahrq.gov/documentation.htm.

The IQIs include the following twenty-eight indicators, which are measured at the provider, or hospital, level:

Volume Indicators	Mortality Indicators for Inpatient Procedures
Esophageal resection volume	Esophageal resection mortality rate
Pancreatic resection volume	Pancreatic resection mortality rate
Abdominal aortic aneurysm (AAA) repair volume	AAA repair mortality rate
Coronary artery bypass graft (CABG) volume	CABG mortality rate
Percutaneous transluminal coronary angioplasty (PTCA) volume	*PTCA mortality rate*[13]
Carotid endarterectomy (CEA) volume	*CEA mortality rate*[5]
	Craniotomy mortality rate
	Hip replacement mortality rate

Mortality Indicators for Inpatient Conditions	Utilization Indicators
Acute myocardial infarction (AMI) mortality rate[14]	Cesarean delivery rate
AMI mortality rate, without transfer cases	Primary Cesarean delivery rate
Congestive heart failure (CHF) mortality rate	Vaginal birth after Cesarean (VBAC) rate[6]
Acute stroke mortality rate	VBAC rate, uncomplicated
Gastrointestinal hemorrhage mortality rate	Laparoscopic cholecystectomy rate
Hip fracture mortality rate	Incidental appendectomy in the elderly rate
Pneumonia mortality rate	Bilateral cardiac catheterization rate

The IQIs also include four area-level utilization indicators that reflect the rate of hospitalization in the area for specific procedures. They are designed using an age- and sex-adjusted population-based denominator and discharge-based numerator. These indicators represent procedures whose use varies widely across relatively similar geographic areas with (in most cases) substantial inappropriate use. The area-level IQIs include the following:

Area-level Utilization Indicators

CABG area rate	Hysterectomy area rate
PTCA area rate	Laminectomy or spinal fusion area rate

A list of each IQI along with the associated reference number, as well as the age of the patient population included in the indicator, is provided in Table 1.

[13] PTCA and CEA mortality are not recommended as standalone indicators, but are suggested as companion measures to the corresponding volume measures.
[14] AMI mortality and VBAC each have two versions: the original AHRQ specification and an alternative specification. See *Inpatient Quality Indicators Technical Specifications* for details.

IQI #3 Pediatric Heart Surgery Volume and IQI #10 Pediatric Heart Surgery Mortality have been moved to the Pediatric Quality Indicators module. All IQIs now apply only to adult populations.

Table 1: Inpatient Quality Indicator (IQI) Variables

Type		IQI number	Indicator	Age categories 18 to 39	40 to 64	65 +
Provider	Volumes	1	Esophageal resection			
		2	Pancreatic resection			
		4	AAA repair			
		5	CABG	No		
		6	PTCA[a]	No		
		7	Carotid endarterectomy			
	Post-procedural mortality Rates	8	Esophageal resection			
		9	Pancreatic resection			
		11	AAA repair			
		12	CABG	No		
		30	*PTCA*[b]	No		
		31	*Carotid endarterectomy*[b]			
		13	Craniotomy			
		14	Hip replacement			
	In-Hospital Mortality rates	15	AMI			
		32	AMI, Without Transfer Cases			
		16	CHF			
		17	Stroke			
		18	GI hemorrhage			
		19	Hip fracture			
		20	Pneumonia			
	Utilization rates	21	Cesarean delivery			
		33	Primary Cesarean delivery			
		22	VBAC (Vaginal Birth After Cesarean), Uncomplicated			
		34	VBAC, All			
		23	Laparoscopic Cholecystectomy			
		24	Incidental appendectomy among elderly	No	No	
		25	Bi-lateral cardiac catheterization			
Area	Utilization rates	26	CABG	No		
		27	PTCA	No		
		28	Hysterectomy			
		29	Laminectomy			

[a] PTCA = percutaneous transluminal coronary angioplasty
[b] PTCA and carotid endarterectomy mortality are not recommended as stand-alone indicators, but are suggested as companion measures to the corresponding volume measures.

1.2 How Can the IQIs be used in Quality Assessment?

The Inpatient Quality Indicators can be used by a variety of players in the health care arena to improve quality of care at the level of individual hospitals, the community, the State, or the nation. The following scenario illustrates one potential application of the IQIs.

A hospital association recognizes its member hospitals' needs for information that can help them evaluate the quality of care they provide. After learning about the IQIs, the association decides to apply the indicators to the discharge abstract data submitted by individual hospitals. For each hospital, the association develops a report with a graphic presentation of the risk-adjusted data to show how that hospital performs on each indicator compared with its peer group, the State as a whole, and other comparable States. National and regional averages are also provided as external benchmarks. Trend data are included to allow the hospital to examine any changing patterns in its performance.

One member hospital, upon receiving the report, convenes an internal work group comprised of both quality improvement professionals and clinicians to review the information and address potential areas for improvements. Since the report is based on administrative data, the work group compares the data with information obtained from other internal sources. For example, to examine the mortality data, they perform chart review for a random sample of patients with a particular condition to verify that the coding is accurate and to ascertain if the death was preventable.

After in-depth analysis of the data and additional chart review, the work group meets with various clinical departments to discuss the results. During those meetings, individual cases are examined and the processes of care are reviewed to identify what patient factors and care processes might have had an impact on patient outcomes. Best practices identified from the literature are also discussed. The work group puts together an internal document that summarizes the findings and makes recommendations for various quality improvement initiatives. The document is shared with the hospital's executives and physician leaders, who strongly support the implementation of several quality improvement projects:

- To improve patient outcomes, the quality improvement team develops and implements comprehensive risk assessment tools and treatment protocols for patients at risk of mortality.
- Physicians refine patient selection criteria for several elective procedures to improve appropriate utilization.
- The hospital reaches out to the local chapter of the American College of Obstetrics and Gynecology and other health care organizations to address the high Cesarean delivery rates among obstetric patients in their community.
- Problems in ICD-9-CM coding are discovered during the chart review process, so health information personnel in the hospital embark on a project to improve communication with physicians to increase the accuracy of coding medical records.

1.3 What Does this Guide Contain?

This guide provides information that hospitals, State data organizations, hospital associations, and others can use to decide how to use the IQIs. First, it describes the origin of the entire family of AHRQ Quality Indicators. Second, it provides an overview of the methods used to identify, select, and evaluate the AHRQ Quality Indicators. Third, the guide summarizes the IQIs specifically, describes strengths and limitations of the indicators, documents the evidence that links the IQIs to the quality of health care services, and then provides in-depth descriptions of each IQI.

A new document, *Inpatient Quality Indicators Technical Specifications* contains the information that was in Appendix A in previous versions. It outlines the specific definitions of each PSI, with complete ICD-9-CM coding specifications. A new section, “Using Different Types of QI Rates” has been added to this Guide. Appendix A now contains links to documents and tools that may be of interest to PSI users.

1.4 Support for Potential and Current Users of the AHRQ QIs

Technical assistance is available, through an electronic user support system monitored by the QI support team, to support users in their application of the IQI software. The same e-mail address may be used to communicate to AHRQ any suggestions for IQI enhancements, general questions, and any QI related comments you may have. AHRQ welcomes your feedback. The Internet address for user support and feedback is: support@qualityindicators.ahrq.gov. AHRQ also offers a listserv to keep you informed on the Quality Indicators (QIs). The listserv is used to announce any QI changes or updates, new tools and resources, and to distribute other QI related information. This is a free service. Sign-up information is available at the QI website at http://www.qualityindicators.ahrq.gov.

2.0 Origins and Background of the Quality Indicators

2.1 Development of the HCUP Quality Indicators

In the early 1990s, in response to requests for assistance from State-level data organizations and hospital associations with inpatient data collection systems, AHRQ developed a set of quality measures that required only the type of information found in routine hospital administrative data—diagnoses and procedures, along with information on patient's age, gender, source of admission, and discharge status. These States were part of the Healthcare Cost and Utilization Project (HCUP), an ongoing Federal-State-private sector collaboration to build uniform databases from administrative hospital-based data collected by State data organizations and hospital associations. Additional information on HCUP is available at the website http://www.ahrq.gov/data/hcup/.

AHRQ developed these measures, called the HCUP Quality Indicators, to take advantage of a readily available data source—administrative data based on hospital claims—and quality measures that had been reported elsewhere.[15] The 33 HCUP QIs included measures for avoidable adverse outcomes, such as in-hospital mortality and complications of procedures; use of specific inpatient procedures thought to be overused, underused, or misused; and ambulatory care sensitive conditions.

Although administrative data cannot provide definitive measures of health care quality, they can be used to provide *indicators* of health care quality that can serve as the starting point for further investigation. The HCUP QIs have been used to assess potential quality-of-care problems and to delineate approaches for dealing with those problems. Hospitals with high rates of poor outcomes on the HCUP QIs have reviewed medical records to verify the presence of those outcomes and to investigate potential quality-of-care problems.[16] For example, one hospital that detected high utilization rates for certain procedures refined patient selection criteria for these procedures to improve appropriate utilization.

2.2 Development of the AHRQ Quality Indicators

Since the original development of the HCUP QIs, the knowledge base on quality indicators has increased significantly. Risk adjustment methods have become more readily available, new measures have been developed, and analytic capacity at the State level has expanded considerably. Based on input from current users and advances to the scientific base for specific indicators, AHRQ funded a project to refine and further develop the original QIs. The project was conducted by the UCSF-Stanford Evidence-Based Practice Center (EPC).

The major constraint placed on the UCSF-Stanford EPC was that the measures could require only the type of information found in hospital discharge abstract data. Further, the data elements required by the measures had to be available from most inpatient administrative data systems. Some State data systems contain innovative data elements, often based on additional information from the medical record. Despite the value of these record-based data elements, the intent of this project was to create measures that were based on a *common denominator discharge data set*, without the need for additional data collection. This was critical for two reasons. First, this constraint would result in a tool that could be used with any inpatient administrative data, thus making it useful to most data systems. Second, this would enable national and regional benchmark rates to be provided using HCUP data, since these benchmark rates would need to be calculated using the universe of data available from the States.

2.3 AHRQ Quality Indicator Modules

[15] Ball JK, Elixhauser A, Johantgen M, et al. *HCUP Quality Indicators, Methods, Version 1.1: Outcome, Utilization, and Access Measures for Quality Improvement*. (AHCPR Publication No. 98-0035). Healthcare Cost and Utilization project (HCUP-3) Research notes: Rockville, MD: Agency for Health Care Policy and Research, 1998.

[16] *Impact: Case Studies Notebook – Documented Impact and Use of AHRQ's Research*. Compiled by Division of Public Affairs, Office of Health Care Information, Agency for Healthcare Research and Quality.

The work of the UCSF-Stanford EPC resulted in the *AHRQ Quality Indicators*, which are available as separate modules:

- **Prevention Quality Indicators**. These indicators consist of “ambulatory care sensitive conditions,” hospital admissions that evidence suggests could have been avoided through high-quality outpatient care or that reflect conditions that could be less severe, if treated early and appropriately.

- **Inpatient Quality Indicators**. These indicators reflect quality of care inside hospitals and include inpatient mortality; utilization of procedures for which there are questions of overuse, underuse, or misuse; and volume of procedures for which there is evidence that a higher volume of procedures is associated with lower mortality.

- **Patient Safety Indicators**. These indicators focus on potentially preventable instances of complications and other iatrogenic events resulting from exposure to the health care system.

- **Pediatric Quality Indicators.** This module, available in February, 2006, contains indicators that apply to the special characteristics of the pediatric population.

The core of the Pediatric Quality Indicators (PDIs) is formed by indicators drawn from the original three modules. Some of these indicators were already geared to the pediatric population (for example, IQI 4 – Pediatric Heart Surgery Volume). These indicators are being removed from the original modules.

Others were adapted from indicators that apply to both adult and pediatric populations. These indicators remain in the original module, but will apply only to adult populations.

3.0 Methods of Identifying, Selecting, and Evaluating the Quality Indicators

In developing the new quality indicators, the UCSF-Stanford EPC applied the Institute of Medicine's widely cited definition of quality care: "the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge."[17] They formulated six specific key questions to guide the development process:

- Which indicators are currently in use or described in the literature that could be *defined using hospital discharge data*?
- What are the *quality relationships* reported in the literature that could be used to define new indicators using hospital discharge data?
- What evidence exists for *indicators not well represented* in the original indicators—pediatric conditions, chronic disease, new technologies, and ambulatory care sensitive conditions?
- Which indicators have *literature-based evidence* to support face validity, precision of measurement, minimum bias, and construct validity of the indicator?
- What *risk-adjustment method* should be suggested for use with the recommended indicators, given the limits of administrative data and other practical concerns?
- Which indicators perform well on *empirical tests* of precision of measurement, minimum bias, and construct validity?

As part of this project, the UCSF-Stanford EPC identified quality indicators reported in the literature and used by health care organizations, evaluated the original quality indicators and potential indicators using literature review and empirical methods, incorporated risk adjustment for comparative analysis, and developed new programs that could be employed by users with their own hospital administrative data. This section outlines the steps used to arrive at a final set of quality measures.

3.1 Step 1: Obtain Background Information on QI Use

The project team at the UCSF-Stanford EPC interviewed 33 individuals affiliated with hospital associations, business coalitions, State data groups, Federal agencies, and academia about various topics related to quality measurement, including indicator use, suggested indicators, and other potential contacts. Interviews were tailored to the specific expertise of interviewees. The sample was not intended to be representative of any population; rather, individuals were selected to include QI users and potential users from a broad spectrum of organizations in both the public and private sectors.

Three broad audiences were considered for the quality measures: health care providers and managers, who could use the quality measures to assist in initiatives to improve quality; public health policy makers, who could use the information from indicators to target public health interventions; and health care purchasers, who could use the measures to guide decisions about health policies.

3.2 Step 2: Search the Literature to Identify Potential QIs

The project team performed a structured review of the literature to identify potential indicators. They used Medline to identify the search strategy that returned a test set of known applicable articles in the most

[17] Institute of Medicine Division of Health Care Services. Medicare: a strategy for quality assurance. Washington, DC: National Academy Press; 1990.

concise manner. Using the Medical Subject Heading (MeSH) terms “Hospital/statistics and numerical data” and “Quality Indicators, Health Care” resulted in approximately 2,600 articles published in 1994 or later. After screening titles and abstracts for relevancy, the search yielded 181 articles that provided information on potential quality indicators based on administrative data.

Clinicians, health services researchers, and other team members abstracted information from these articles in two stages. In the first stage, preliminary abstraction, they evaluated each of the 181 identified articles for the presence of a defined quality indicator, clinical rationale, and strengths and weaknesses. To qualify for full abstraction, the articles must have explicitly defined a novel quality indicator. Only 27 articles met this criterion. The team collected information on the definition of the quality indicator, validation, and rationale during full abstraction.

In addition, they identified additional potential indicators using the CONQUEST database; the National Library of Healthcare Indicators developed by the Joint Commission on Accreditation of Healthcare Organizations (JCAHO); a list of ORYX-approved indicators provided by JCAHO; and telephone interviews.

3.3 Step 3: Review the Literature to Evaluate the QIs According to Predetermined Criteria

The project team evaluated each potential quality indicator against the following six criteria, which were considered essential for determining the reliability and validity of a quality indicator:

- **Face validity.** An adequate quality indicator must have sound clinical or empirical rationale for its use. It should measure an important aspect of quality that is subject to provider or health care system control.

- **Precision.** An adequate quality indicator should have relatively large variation among providers or areas that is not due to random variation or patient characteristics. This criterion measures the impact of chance on apparent provider or community health system performance.

- **Minimum bias.** The indicator should not be affected by systematic differences in patient case-mix, including disease severity and comorbidity. In cases where such systematic differences exist, an adequate risk adjustment system should be possible using available data.

- **Construct validity.** The indicator should be related to other indicators or measures intended to measure the same or related aspects of quality. For example, improved performance on measures of inpatient care (such as adherence to specific evidence-based treatment guidelines) ought to be associated with reduced patient complication rates.

- **Fosters real quality improvement.** The indicator should be robust to possible provider manipulation of the system. In other words, the indicator should be insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care.

- **Application.** The indicator should have been used in the past or have high potential for working well with other indicators. Sometimes looking at groups of indicators together is likely to provide a more complete picture of quality.

Based on the initial review, the team identified and evaluated over 200 potential indicators using these criteria. Of this initial set, 45 indicators passed this initial screen and received comprehensive literature and empirical evaluation. In some cases, whether an indicator complemented other promising indicators was a consideration in retaining it, allowing the indicators to provide more depth in specific areas.

For this final set of 45 indicators, the team reviewed an additional 2,000 articles to provide evidence on indicators during the evaluation phase. They searched Medline for articles relating to each of the six areas of evaluation described above. Clinicians and health services researchers reviewed the literature for evidence and prepared a referenced summary description on each indicator.

As part of the review process, the team assessed the link between each indicator and health care quality along the following dimensions:

- **Proxy.** Some indicators do not specifically measure a patient outcome or a process measure of quality. Rather, they measure an aspect of care that is correlated with process measures of quality or patient outcomes. These indicators are best used in conjunction with other indicators measuring similar aspects of clinical care, or when followed with more direct and in-depth investigations of quality.

- **Selection bias.** Selection bias results when a substantial percentage of care for a condition is provided in the outpatient setting, so the subset of inpatient cases may be unrepresentative. In these cases, examination of outpatient care or emergency room data may help reduce selection bias.

- **Information bias.** Quality indicators are based on information available in hospital discharge data sets, but some missing information may actually be important to evaluating the outcomes of hospital care. In these cases, examination of missing information may help to improve indicator performance.

- **Confounding bias.** Patient characteristics may substantially affect performance on a measure and may vary systematically across areas. In these cases, adequate risk adjustment may help to improve indicator performance.

- **Unclear construct validity.** Problems with construct validity include uncertain or poor correlations with widely accepted process measures or with risk-adjusted outcome measures. These indicators would benefit from further research to establish their relationship with quality care.

- **Easily manipulated.** Quality indicators may create perverse incentives to improve performance without actually improving quality. Although very few of these perverse responses have been proven, they are theoretically important and should be monitored to ensure true quality improvement.

- **Unclear benchmark.** For some indicators, the “right rate” has not been established, so comparison with national, regional, or peer group means may be the best benchmark available. Very low IQI rates may flag an underuse problem, that is, providers may fail to hospitalize patients who would benefit from inpatient care. On the other hand, overuse of acute care resources may potentially occur when patients who do not clinically require inpatient care are hospitalized.

3.4 Step 4: Perform a Comprehensive Evaluation of Risk Adjustment

The project team identified potential risk-adjustment systems by reviewing the applicable literature and asking the interviewees in step 1 to identify their preferences. Generally, users preferred that the system be (1) open, with published logic; (2) cost-effective, with data collection costs minimized and additional data collection being well justified; (3) designed using a multiple-use coding system, such as those used for reimbursement; and (4) officially recognized by government, hospital groups, or other organizations.

Although no severity adjustment system based solely on administrative data is superior for all purposes, risk adjustment systems based on diagnosis-related groups (DRGs) seemed to meet the criteria for this evaluation better than other alternatives. Specifically, it was presumed that because a DRG-based system relies on the same diagnostic groups used for reimbursement, there may be more accurate coding as a result of the financial and audit incentives associated with use of DRGs.

One DRG-based system in particular—all-patient refined (APR)-DRGs—appeared to be promising for several reasons. First, APR-DRGs are based on a refinement of two previously developed systems (R-DRGs and AP-DRGs) and take advantage of the strengths of both of these systems. Second, APR-DRGs were enhanced to provide improved risk adjustment for pediatric cases; to take advantage of information on comorbidities and non-operating room procedures; and to allow the interaction of secondary diagnoses, principal diagnosis, and age to influence the assignment of severity classes. Third, APR-DRGs have been reported to perform well in predicting resource use and death when compared to other DRG-based systems. Fourth, APR-DRGs have been used with "smoothing" techniques, the statistical methods incorporated into the QI software, thus compatibility with the QI software was ensured. A majority of the users interviewed already used the 3M™ All-Patient Refined DRG[18] (APR-DRG) system, which has been reported to perform well in predicting resource use and death when compared to other DRG-based systems. Even though the system is proprietary, the burden on the group of potential QI users would be smaller than with another system that was less widely employed.

APR-DRGs were used to conduct indicator evaluations to determine the impact of measured differences in patient severity on the relative performance of providers and to provide the basis for implementing APR-DRGs as an optional risk-adjustment system for hospital-level QI measures. The implementation of APR-DRGs is based on an ordinary least squares regression model. Area indicators were risk-adjusted only for age and sex differences.

3.5 Step 5: Evaluate the Indicators Using Empirical Analyses

The project team conducted extensive empirical testing of all potential indicators using the 1995-97 HCUP State Inpatient Databases (SID) and Nationwide Inpatient Sample (NIS) to determine precision, bias, and construct validity. The 1997 SID contains uniform data on inpatient stays in community hospitals for 22 States covering approximately 60% of all U.S. hospital discharges. The NIS is designed to approximate a 20% sample of U.S. community hospitals and includes all stays in the sampled hospitals. Each year of the NIS contains between 6 million and 7 million records from about 1,000 hospitals. The NIS combines a subset of the SID data, hospital-level variables, and hospital and discharge weights for producing national estimates. The project team conducted tests to examine three things: precision, bias, and construct validity.

Precision. The first step in the analysis involved precision tests to determine the reliability of the indicator for distinguishing real differences in provider performance. For indicators that may be used for quality improvement, it is important to know with what precision, or surety, a measure can be attributed to an actual construct rather than random variation.

For each indicator, the variance can be broken down into three components: variation within a provider (actual differences in performance due to differing patient characteristics), variation among providers (actual differences in performance among providers), and random variation. An ideal indicator would have a substantial amount of the variance explained by between-provider variance, possibly resulting from differences in quality of care, and a minimum amount of random variation. The project team performed four tests of precision to estimate the magnitude of between-provider variance on each indicator:

- Signal standard deviation was used to measure the extent to which performance of the QI varies systematically across hospitals or areas.
- Provider/area variation share was used to calculate the percentage of signal (or true) variance relative to the total variance of the QI.

[18] Information on the 3M™ APR-DRG system is available at http://www.3m.com/us/healthcare/his/products/coding/refined_drg.jhtml.

- Signal-to-noise ratio was used to measure the percentage of the apparent variation in QIs across providers that is truly related to systematic differences across providers and not random variations (noise) from year to year.

- In-sample R-squared was used to identify the incremental benefit of applying multivariate signal extraction methods for identifying additional signal on top of the signal-to-noise ratio.

In general, random variation is most problematic when there are relatively few observations per provider, when adverse outcome rates are relatively low, and when providers have little control over patient outcomes or variation in important processes of care is minimal. If a large number of patient factors that are difficult to observe influence whether or not a patient has an adverse outcome, it may be difficult to separate the "quality signal" from the surrounding noise. Two signal extraction techniques were applied to improve the precision of an indicator:

- Univariate methods were used to estimate the "true" quality signal of an indicator based on information from the specific indicator and 1 year of data.

- Multivariate signal extraction (MSX) methods were used to estimate the "true" quality signal based on information from a set of indicators and multiple years of data. In most cases, MSX methods extracted additional signal, which provided much more precise estimates of true hospital or area quality.

Bias. To determine the sensitivity of potential QIs to bias from differences in patient severity, unadjusted performance measures for specific hospitals were compared with performance measures that had been adjusted for age and gender. All of the Prevention QIs and some of the IQIs could only be risk-adjusted for age and sex. The 3M APR-DRG System Version 12 with Severity of Illness and Risk of Mortality subclasses was used for risk adjustment of the utilization indicators and the in-hospital mortality indicators, respectively. Five empirical tests were performed to investigate the degree of bias in an indicator:

- Rank correlation coefficient of the area or hospital with (and without) risk adjustment—gives the overall impact of risk adjustment on relative provider or area performance.

- Average absolute value of change relative to mean—highlights the amount of absolute change in performance, without reference to other providers' performance.

- Percentage of highly ranked hospitals that remain in high decile—reports the percentage of hospitals or areas that are in the highest deciles without risk adjustment that remain there after risk adjustment is performed.

- Percentage of lowly ranked hospitals that remain in low decile—reports the percentage of hospitals or areas that are in the lowest deciles without risk adjustment that remain there after risk adjustment is performed.

- Percentage that change more than two deciles—identifies the percentage of hospitals whose relative rank changes by a substantial percentage (more than 20%) with and without risk adjustment.

Construct validity. Construct validity analyses provided information regarding the relatedness or independence of the indicators. If quality indicators do indeed measure quality, then two measures of the same construct would be expected to yield similar results. The team used factor analysis to reveal underlying patterns among large numbers of variables—in this case, to measure the degree of relatedness between indicators. In addition, they analyzed correlation matrices for indicators.

4.0 Summary Evidence on the Inpatient Quality Indicators

The rigorous evaluations performed by the UCSF-Stanford EPC, based on literature review and empirical testing of indicators, resulted in 29 indicators that reflect inpatient volume, mortality, and utilization. (Two additional mortality indicators are provided that are recommended for use only with the corresponding volume measures.) IQI Version 1.2, Revision 3, included three additional measures—AMI Mortality without transfer cases, VBAC rate uncomplicated, and an indicator for Primary Cesarean delivery rate. Five of the provider-level IQIs and three area-level IQIs were included in the original HCUP QIs—Cesarean delivery rate, incidental appendectomy in the elderly rate, VBAC rate, laparoscopic cholecystectomy rate, hip replacement mortality rate, CABG area rate, hysterectomy area rate, and laminectomy or spinal fusion area rate.

4.1 Version 3.0 Inpatient Quality Indicators

A modified version of the process described in Section 3.0 is repeated on an annual basis when the IQIs are evaluated and new indicators are considered. With this release two of the original indicators dealing with pediatric heart surgery have been moved to the PDIs.

New micropolitan statistical areas and updated metropolitan statistical areas were established by the federal Office of Management and Budget (OMB) circular 03-04 (last revised December 4, 2005). To reflect these changes, all IQI documentation now refers to Metro Area instead of MSA. The SAS and SPSS software allows users to specify stratification by county level with U.S. Census FIPS or modified FIPS, or by Metro Area with OMB 1999 or OMB 2003 definition. The AHRQ QI Windows Application allows users to generate reports stratified by all four of these, as well as by State.

Table 2 summarizes the results of the literature review and empirical evaluations on the IQIs. The table lists each indicator, provides its definition, rates its empirical performance, recommends a risk adjustment strategy, and summarizes important caveats identified from the literature review.

Rating of performance on empirical evaluations, as described in Step 5 in section 3.5, ranged from 0 to 26. (The average score for the mortality IQIs is 6.2; the average score for the utilization IQIs is 19.3.) The scores were intended as a guide for summarizing the performance of each indicator on four empirical tests of precision (signal variance, area-level share, signal ratio, and R-squared) and five tests of minimum bias (rank correlation, top and bottom decile movement, absolute change, and change over two deciles), as described in the previous section.

The magnitude of the scores, shown in the Empirical Performance column, provides an indication of the relative rankings of the indicators. These scores were based on indicator performance after risk-adjustment and smoothing, that is, they represent the "best estimate" of the indicator's true value after accounting for case-mix and reliability. The score for each individual test is an ordinal ranking (e.g., very high, high, moderate, and low). The final summary score was derived by assigning a weight to each ranking (e.g., 3, 2, 1, 0) and summing across these nine individual tests. Higher scores indicate better performance on the empirical tests.

The Literature Review Caveats column summarizes evidence specific to each potential concern on the link between the IQIs and quality of care, as described in step 3 above. A question mark (**?**) indicates that the concern is theoretical or suggested, but no specific evidence was found in the literature. A check mark (✔) indicates that the concern has been demonstrated in the literature. For additional details on the results of the literature review, see "Detailed Evidence for the Inpatient Quality Indicators."

A complete description of each IQI is included in Section 5.0 "Detailed Evidence for Inpatient Quality Indicators" and in the document *Inpatient Quality Indicators Technical Specifications.* See Appendix A for links to additional information.

Table 2: AHRQ Inpatient Quality Indicators Empirical Evaluations

Indicator Name (Number)	Description	Risk Adjustment Used by QI Software	Empirical Performance[a]	Literature Review Caveats[b]
Volume Indicators				
Esophageal Resection Volume (IQI 1)	Raw volume compared to annual thresholds (6 and 7 procedures).	Not applicable.	Avg. Volume = 4.39 Avg. Volume SD = 9.41 Rating = Not applicable	✔ Proxy ? Easily manipulated
Pancreatic Resection Volume (IQI 2)	Raw volume compared to annual thresholds (10 and 11 procedures).	Not applicable.	Avg. Volume = 5.99 Avg. Volume SD = 12.32 Rating = Not applicable	✔ Proxy ? Easily manipulated
Abdominal Aortic Aneurysm Repair (AAA) Volume (IQI 4)	Raw volume compared to annual thresholds (10 and 32 procedures).	Not applicable.	Avg. Volume = 17.63 Avg. Volume SD = 25.23 Rating = Not applicable	✔ Proxy ? Easily manipulated
Coronary Artery Bypass Graft (CABG) Volume (IQI 5)	Raw volume compared to annual thresholds (100 and 200 procedures).	Not applicable.	Avg. Volume = 294.71.59 Avg. Volume SD = 260.39 Rating = Not applicable	✔ Proxy ? Easily manipulated
Percutaneous Transluminal Coronary Angioplasty (PTCA) Volume (IQI 6)	Raw volume compared to annual thresholds (200 and 400 procedures).	Not applicable.	Avg. Volume = 535.05 Avg. Volume SD = 593.86 Rating = Not applicable	✔ Proxy ? Selection bias ✔ Easily manipulated
Carotid Endarterectomy (CEA) Volume (IQI 7)	Raw volume compared to annual thresholds (50 and 101 procedures).	Not applicable.	Avg. Volume = 54.84 Avg. Volume SD =59.62 Rating = Not applicable	✔ Proxy ✔ Easily manipulated
Mortality Indicators for Inpatient Procedures				
Esophageal Resection Mortality Rate (IQI 8)	Number of deaths per 100 esophageal resections for cancer.	APR-DRG, though impact may be impaired by skewed distribution.	Provider Rate = 10.86 Provider SD = 26.80 Pop. Rate = 8.54 Rating = 8	? Confounding bias ? Unclear construct validity
Pancreatic Resection Mortality Rate (IQI 9)	Number of deaths per 100 pancreatic resections for cancer.	APR-DRG, though impact may be impaired by skewed distribution.	Provider Rate = 10.20 Provider SD = 24.49 Pop. Rate = 6.60 Rating = 5	? Confounding bias ? Unclear construct validity

Indicator Name (Number)	**Description**	**Risk Adjustment Used by QI Software**	**Empirical Performance**[a]	**Literature Review Caveats**[b]
AAA Repair Mortality Rate (IQI 11)	Number of deaths per 100 AAA repairs.	APR-DRG, though impact may be impaired by skewed distribution.	Provider Rate = 13.11 Provider SD = 21.92 Pop. Rate = 7.27 Rating = 8	✔ Confounding bias ? Unclear construct validity
CABG Mortality Rate (IQI 12)	Number of deaths per 100 CABG procedures.	APR-DRG.	Provider Rate = 3.59 Provider SD = 3.99 Pop. Rate = 3.39 Rating = 5	? Selection bias ✔ Confounding bias ? Unclear construct validity ? Easily manipulated
PTCA Mortality Rate[c] *(IQI 30)*	*Number of deaths per 100 PTCAs*	*APR-DRG.*	*Provider Rate = 1.92* *Provider SD = 6.04* *Pop. Rate = 1.30* *Rating = not available*	*Not evaluated during initial literature review*
CEA Mortality Rate[c] *(IQI 31)*	*Number of deaths per 100 CEAs.*	*APR-DRG.*	*Provider Rate = 0.75* *Provider SD = 3.80* *Pop. .Rate = 0.62* *Rating = not available*	*Not evaluated during initial literature review*
Craniotomy Mortality Rate (IQI 13)	Number of deaths per 100 craniotomies.	APR-DRG.	Provider Rate = 8.82 Provider SD = 11.01 Pop. Rate = 7.10 Rating = 6	✔ Confounding bias ? Unclear construct validity
Hip replacement mortality rate (IQI 14)	Number of deaths per 100 hip replacements.	APR-DRG.	Provider Rate = 0.47 Provider SD = 2.94 Pop. Rate = 0.29 Rating = 3	? Selection bias ? Confounding bias ? Unclear construct validity
Mortality Indicators for Inpatient Conditions				
Acute Myocardial Infarction (AMI) Mortality Rate (IQI 15)	Number of deaths per 100 discharges for AMI.	APR-DRG.	Provider Rate = 14.79 Provider SD = 14.35 Pop. Rate = 8.85 Rating = 5	✔ Information bias ✔ Confounding bias
Acute Myocardial Infarction (AMI) Mortality Rate, Without Transfer Cases (IQI 32)	Number of deaths per 100 discharges for AMI.	APR-DRG.	Provider Rate = 14.99 Provider SD = 13.93 Pop. Rate = 9.75. *Rating = not available*	Not evaluated during initial literature review
Congestive Heart Failure (CHF) Mortality Rate (IQI 16)	Number of deaths per 100 discharges for CHF.	APR-DRG.	Provider Rate = 5.25 Provider SD = 7.86 Pop. Rate = 4.33 Rating = 6	✔ Selection bias ✔ Information bias ✔ Confounding bias
Acute Stroke Mortality Rate (IQI 17)	Number of deaths per 100 discharges for stroke.	APR-DRG	Provider Rate = 10.57 Provider SD = 10.33 Pop. Rate = 11.16 Rating = 10	✔ Selection bias ? Information bias ✔ Confounding bias

Indicator Name (Number)	Description	Risk Adjustment Used by QI Software	Empirical Performance[a]	Literature Review Caveats[b]
Gastrointestinal (GI) Hemorrhage Mortality Rate (IQI 18)	Number of deaths per 100 discharges for GI hemorrhage.	APR-DRG.	Provider Rate = 3.29 Provider SD = 7.16 Pop. Rate = 3.00 Rating = 5	✔ Confounding bias ? Unclear construct validity
Hip fracture Mortality Rate (IQI 19)	Number of deaths per 100 discharges for hip fracture.	APR-DRG.	Provider Rate = 3.80 Provider SD = 8.20 Pop. Rate = 3.18 Rating = 10	? Information bias ✔ Confounding bias ? Unclear construct validity
Pneumonia Mortality Rate (IQI 20)	Number of deaths per 100 discharges for pneumonia.	APR-DRG.	Provider Rate = 7.53 Provider SD = 6.31 Pop. Rate = 7.75 Rating = 7	✔ Selection bias ? Information bias ✔ Confounding bias
Utilization Indicators - Provider (Hospital) Level				
Cesarean Delivery Rate (IQI 21)	Number of Cesarean deliveries per 100 deliveries.	Age.	Provider Rate = 24.48 Provider SD = 8.85 Pop. Rate = 24.47 Rating = 17	? Confounding bias ? Unclear construct validity ? Unclear benchmark
Primary Cesarean Delivery Rate (IQI 33)	Number of Cesarean deliveries per 100 deliveries in women with no history of previous Cesarean delivery.	Age.	Provider Rate = 15.17 Provider SD = 7.05 Pop. Rate = 15.26 *Rating = not available*	Not evaluated during initial literature review
Vaginal Birth After Cesarean (VBAC) Rate, Uncomplicated (IQI 22)	Number of vaginal births per 100 deliveries in women with previous Cesarean delivery.	Age.	Provider Rate = 13.31 Provider SD = 11.68 Pop. Rate = 15.30 Rating = 19	✔ Selection bias ? Confounding bias ? Unclear construct validity ? Unclear benchmark
Vaginal Birth After Cesarean (VBAC) Rate, All (IQI 34)	Number of vaginal births per 100 deliveries in women with history of previous Cesarean delivery.	Age.	Provider Rate = 12.91 Provider SD = 11.22 Pop. Rate = 14.84 *Rating = not available*	Not evaluated during initial literature review
Laparoscopic Cholecystectomy Rate (IQI 23)	Number of laparoscopic cholecystectomies per 100 cholecystectomies.	Age and sex.	Provider Rate = 74.74 Provider SD = 19.55 Pop. Rate = 75.55 Rating = 20	✔ Selection bias ✔ Confounding bias ? Unclear construct validity ✔ Easily manipulated ✔ Unclear benchmark
Incidental Appendectomy in the Elderly Rate (IQI 24)	Number of incidental appendectomies per 100 abdominal surgeries.	APR-DRG.	Provider Rate = 2.50 Provider SD = 4.57 Pop. Rate = 2.30 Rating = 13	? Unclear construct validity ? Easily manipulated
Bilateral Cardiac Catheterization Rate (IQI 25)	Number of bilateral catheterizations per 100 cardiac catheterizations.	APR-DRG.	Provider Rate = 8.45 Provider SD = 12.25 Pop. Rate = 7.13 Rating = 25	? Selection bias ? Unclear construct validity

Indicator Name (Number)	Description	Risk Adjustment Used by QI Software	Empirical Performance[a]	Literature Review Caveats[b]
Utilization Indicators - Area Level				
CABG Rate[d] (IQI 26)	Number of CABGs per 100,000 population.	Age and sex.	Area Rate = 278.82 Area SD = 129.69 Pop. Rate = 241.41 Rating = 19	✔ Proxy ✔ Unclear construct validity ✔ Unclear benchmark
PTCA Rate[d] (IQI 27)	Number of PTCAs per 100,000 population.	Age and sex.	Area Rate = 619.42 Area SD = 277.13 Pop. Rate = 568.29 Rating = 19	✔ Proxy ? Selection bias ✔ Unclear construct validity ✔ Unclear benchmark
Hysterectomy Rate (IQI 28)	Number of hysterectomies per 100,000 population.	Age and additional factors such as parity.	Area Rate = 560.81 Area SD = 205.4 Pop. Rate = 464.34 Rating = 22	✔ Proxy ? Confounding bias ✔ Unclear construct validity ✔ Unclear benchmark
Laminectomy or Spinal Fusion Rate (IQI 29)	Number of laminectomies per 100,000 population.	Age and sex.	Area Rate = 296.91 Area SD = 143.56 Pop. Rate = 252.77 Rating = 20	✔ Proxy ✔ Unclear construct validity ✔ Unclear benchmark

[a] Notes under **Empirical Performance**:
Provider Rates – The national observed (unadjusted) and unweighted rates for providers (hospitals) and their standard deviations (SD) were calculated using the HCUP Year 2003 SID from 38 states. Provider rates are per 100 and were based on 4,688 providers.
Area Rates – The national observed (unadjusted) and unweighted rates for areas (counties) and their standard deviations (SD) were based on 2,570 geographic areas (counties) in the HCUP Year 2003 SID from 38 states. Area rates are per 100,000.
Population Rates - The population rates are weighted provider and area rates (weighted by the number of discharges for each indicator or area populations).
Ratings - Higher ratings in the Empirical Performance column indicate better performance on the nine empirical tests. .

[b] Notes under **Literature Review Caveats**:
Proxy – Indicator does not directly measure patient outcomes but an aspect of care that is associated with the outcome; thus, it is best used with other indicators that measure similar aspects of care.
Confounding bias – Patient characteristics may substantially affect the performance of the indicator; risk adjustment is recommended.
Unclear construct – There is uncertainty or poor correlation with widely accepted process measures.
Easily manipulated – Use of the indicator may create perverse incentives to improve performance on the indicator without truly improving quality of care.
Unclear benchmark – The "correct rate" has not been established for the indicator; national, regional, or peer group averages may be the best benchmark available.
? – The concern is theoretical or suggested, but no specific evidence was found in the literature.
✔ – Indicates that the concern has been demonstrated in the literature.

[c] PTCA and CEA mortality are not recommended as stand-alone indicators, but are suggested as companion measures to the corresponding volume measures.

[d] CABG and PTCA area utilization are not recommended as stand-alone indicators. They are designed only for use with the corresponding volume and/or mortality measures.

4.2 Strengths and Limitations in Using the IQIs

This collection of AHRQ Quality Indicators represents the current state-of-the-art in assessing quality of care using hospital administrative data. However, these indicators must be used cautiously, because the administrative data on which the indicators are based are not collected for research purposes or for measuring quality of care, but for billing purposes. While these data are relatively inexpensive and convenient to use—and represent a rich data source that can provide valuable information—they should not be used as a definitive source of information on quality of health care. At least three limitations of administrative data warrant caution:

- Coding differences across hospitals. Some hospitals code more thoroughly than others, making “fair” comparisons across hospitals difficult.

- Ambiguity about when a condition occurs. Most administrative data cannot distinguish unambiguously whether a specific condition was present at admission or whether it occurred during the stay (i.e., a possible complication).

- Limitations in ICD-9-CM coding. The codes themselves are often not specific enough to adequately characterize a patient’s condition, which makes it impossible to perfectly risk-adjust any administrative data set, thus fair comparisons across hospitals become difficult.

Ideally, the results on AHRQ IQIs for individual hospitals should be made available to those hospitals, with information on averages for a peer group, for the State, and for the nation. This information can be used by individual hospitals to launch investigations into reasons for potential quality problems. Further study may:

- Reveal real quality problems for which quality improvement programs can be initiated.

- Uncover problems in data collection that can be remedied through stepped-up efforts to code more diligently.

- Determine that additional clinical information is required to understand the quality issues, beyond what can be obtained through billing data alone.

In short, the AHRQ IQIs are a valuable tool that takes advantage of readily available data to flag potential quality-of-care problems.

4.3 Questions for Future Work

The limitations discussed above suggest some directions for future work on development and use of the IQIs. Additional data and linkages could provide insights into whether the findings represent true quality problems, and could facilitate the exploration of potential interventions to prevent such events.

- Hospitals with higher than average mortality rates for specific procedures or conditions should probe the underlying reasons: Are patients more severely ill? Is there a problem in the selection of patients for this particular procedure? Is there a quality-of-care problem? Although the mortality indicators use APR-DRG risk adjustment, limitations in the clinical sensitivity of administrative data mean that it is not possible to unambiguously measure and control for patient severity of illness. These indicators provide a starting point for further investigations that might explore severity of illness differences.

- For hospitals with low volumes of particular procedures, how do patients fare? What is the mortality rate for patients who receive this procedure at this hospital compared with other hospitals? What is the resource use associated with receiving this procedure at this hospital compared with other hospitals? Is there evidence of higher complication rates that suggest a problem in quality of care?

- What are potential explanations for hospitals with higher-than-average utilization rates? Is this hospital a referral center for this procedure? Do patients come from outside the area to receive their procedures at this hospital? Or is there evidence that patients from this area are receiving a greater number of procedures than expected? The AHRQ area-level IQIs use either the county (Metro Area) where the hospital is located or the county (Metro Area) of the patient's residence to define areas. The default is the hospital location because the IQIs presume the common denominator discharge data set (data elements routinely available across most discharge data systems); information such as the patient's county of residence is often not available. High area rates might be due to patients admitted to a hospital that live outside of the county where the hospital is located. The Metro Area option is an alternative (patients admitted to a hospital are less likely to live outside the hospital's Metro Area). The preferred option is to use the county (Metro Area) of the residence of the patient. Then the area rate reflects the number of admissions for residents of that area to any hospital, regardless of location.

- For two indicators, bilateral cardiac catheterization and incidental appendectomy, very few, if any, of there procedures are expected. Records for these patients could be examined to discern a possible justification for performing these procedures.

5.0 Detailed Evidence for Inpatient Quality Indicators

This section provides an abbreviated presentation of the details of the literature review and the empirical evaluation for each IQI, including:

- The relationship between the indicator and quality of health care services
- A suggested benchmark or comparison
- The definition of each indicator
- The numerator (or outcome of interest)
- The denominator (or population at risk)
- The results of the empirical testing

The descriptions for each indicator include a discussion of the summary of evidence, the limitations on using each indicator, and details on the following:

- Face validity – Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?
- Precision – Is there a substantial amount of provider or community level variation that is not attributable to random variation?
- Minimum bias – Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?
- Construct validity – Does the indicator perform well in identifying true (or actual) quality of care problems?
- Fosters true quality improvement – Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?
- Prior use – Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

Population rates based on all eligible discharges are calculated using 2003 SID from 38 states.[19] These rates are also reported in Table 2.

[19] The state data organizations that participated in the 2003 HCUP SID: Arizona Department of Health Services; California Office of Statewide Health Planning & Development; Colorado Health & Hospital Association; Connecticut - Chime, Inc.; Florida Agency for Health Care Administration; Georgia: An Association of Hospitals & Health Systems; Hawaii Health Information Corporation; Illinois Health Care Cost Containment Council; Indiana Hospital & Health Association; Iowa Hospital Association; Kansas Hospital Association; Kentucky Department for Public Health; Maine Health Data Organization; Maryland Health Services Cost Review; Massachusetts Division of Health Care Finance and Policy; Michigan Health & Hospital Association; Minnesota Hospital Association; Missouri Hospital Industry Data Institute; Nebraska Hospital Association; Nevada Department of Human Resources; New Hampshire Department of Health & Human Services; New Jersey Department of Health & Senior Services; New York State Department of Health; North Carolina Department of Health and Human Services; Ohio Hospital Association; Oregon Association of Hospitals & Health Systems; Pennsylvania Health Care Cost Containment Council; Rhode Island Department of Health; South Carolina State Budget & Control Board; South Dakota Association of Healthcare Organizations; Tennessee Hospital Association; Texas Health Care Information Council; Utah Department of Health; Vermont Association of Hospitals and Health Systems; Virginia Health Information; Washington State Department of Health; West Virginia Health Care Authority; Wisconsin Department of Health & Family Services.

A full report on the literature review and empirical evaluation can be found in *Refinement of the HCUP Quality Indicators*. Detailed coding information for each IQI, previously contained in this document, is now in a separate document *Inpatient Quality Indicators Technical Specifications.* See Appendix A for links to these documents.

5.1 Esophageal Resection Volume (IQI 1)

Esophageal surgery is a rare procedure that requires technical proficiency; and errors in surgical technique or management may lead to clinically significant complications, such as sepsis, pneumonia, anastomotic breakdown, and death.

Relationship to Quality	Higher volumes have been associated with better outcomes, which represent better quality.
Benchmark	Threshold 1: 6 or more procedures per year Threshold 2: 7 or more procedures per year [21]
Definition	Raw volume of provider-level esophageal resection.
Numerator	Discharges with ICD-9-CM codes of 424x, 425x or 426x in any procedure field. Age 18 years and older. Exclude cases: • MDC 14 (pregnancy, childbirth, and puerperium) • MDC 15 (newborns and other neonates)
Denominator	Not applicable.
Type of Indicator	Provider Level, Procedure Volume Indicator
Empirical Rating	Not applicable.

Summary of Evidence

The relative rarity of esophageal resection results in an indicator that is less precise than most volume indicators, although still highly adequate for use as a quality indicator. Hospitals should examine more than one year of data if possible and average volumes for a more precise estimate. Hospitals may also consider use with the pancreatic resection indicator, another complex cancer surgery. The volume-outcome relationship on which this indicator is based may not hold over time, as providers become more experienced or as technology changes.

Most hospitals perform fewer than 10 procedures in a 5-year period; however, relatively strong relationships between volume and outcome—specifically post-operative mortality—have been noted in the literature.

Empirical evidence shows that a low percentage of procedures were performed at high-volume hospitals. At threshold 1, 39.5% of esophageal resection procedures were performed at high-volume providers (and 8.6% of providers are high volume).[20] At threshold 2, 34.3% were performed at high-volume providers (and 6.4% of providers are high volume).[21 22]

Limitations on Use

As a volume indicator, esophageal resection is a proxy measure for quality and should be used with other indicators.

Details

Face validity: Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?

The face validity of esophageal resection depends on whether a strong association with outcomes of care is both plausible and widely accepted in the professional community. No consensus recommendations regarding minimum procedure volume currently exist.

Precision : Is there a substantial amount of provider or community level variation that is not attributable to random variation?

---

[20]Patti MG, Corvera CU, Glasgow RE, et al. A hospital's annual rate of esophagectomy influences the operative mortality rate. J Gastrointest Surg 1998;2(2):186-92.

[21]Dudley RA, Johansen KL, Brand R, et al. Selective referral to high-volume hospitals: estimating potentially avoidable deaths. JAMA 2000;283(9):1159-66.

[22]Nationwide Inpatient Sample and State Inpatient Databases. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/data/hcup

Esophageal resection is measured accurately with discharge data. Most facilities perform 10 or fewer esophagectomies for cancer during a 5-year period; therefore, this indicator is expected to have poor precision.

Minimal bias: Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?

Risk adjustment is not appropriate, because volume measures are not subject to bias due to disease severity and comorbidities.

Construct validity: Does the indicator perform well in identifying true (or actual) quality of care problems?

Higher volumes have been repeatedly associated with better outcomes after esophageal surgery, although these findings may be limited by inadequate risk adjustment of the outcome measure.

Only one study used clinical data to estimate the association between hospital volume and mortality following esophageal cancer surgery. Begg et al. analyzed retrospective data from the Surveillance, Epidemiology, and End Results (SEER)-Medicare linked database from 1984 through 1993.[23] The crude 30-day mortality rate was 17.3% at hospitals that performed 1-5 esophagectomies on Medicare patients during the study period, versus 3.9% and 3.4% at hospitals that performed 6-10 and 11 or more esophagectomies, respectively. The association between volume and mortality remained highly significant (p<.001) in a multivariate model, adjusting for the number of comorbidities, cancer stage and volume, and age.

Studies based on California and Maryland data found that the risk-adjusted mortality rates at low-volume hospitals were around 3.0 times those at high-volume hospitals.[24] [25]

Empirical evidence shows that esophageal resection volume—after adjusting for age, sex, and APR-DRG—is moderately and negatively correlated with mortality for esophageal resection (r=-.29, p<.05), as well as mortality after other cancer resection procedures.[26]

Fosters true quality improvement: Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?

Low-volume providers may attempt to increase their volume without improving quality of care by performing the procedure on patients who may not qualify or benefit from the procedure. Additionally, shifting procedures to high-volume providers may impair access to care for certain types of patients.

Prior use: Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

Esophageal cancer surgical volume has not been widely used as an indicator of quality.

[23]Begg CB, Cramer LD, Hoskins WJ, et al. Impact of hospital volume on operative mortality for major cancer surgery. JAMA 1998;280(20):1747-51.
[24]Patti MG, Corvera CU, Glasgow RE, et al. A hospital's annual rate of esophagectomy influences the operative mortality rate. J Gastrointest Surg 1998;2(2):186-92.
[25]Gordan 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.
[26]Nationwide Inpatient Sample.

5.2 Pancreatic Resection Volume (IQI 2)

Pancreatic resection is a rare procedure that requires technical proficiency; and errors in surgical technique or management may lead to clinically significant complications, such as sepsis, anastomotic breakdown, and death.

Relationship to Quality	Higher volumes have been associated with better outcomes, which represent better quality.
Benchmark	Threshold 1: 10 or more procedures per year Threshold 2: 11 or more procedures per year [28]
Definition	Raw volume of provider-level pancreatic resection.
Numerator	Discharges with ICD-9-CM codes of 526 or 527 in any procedure field. Age 18 years and older. Exclude cases: • MDC 14 (pregnancy, childbirth, and puerperium) • MDC 15 (newborns and other neonates)
Denominator	Not applicable.
Type of Indicator	Provider Level, Procedure Volume Indicator
Empirical Rating	Not applicable.

Summary of Evidence

The relative rarity of pancreatic resection results in an indicator that is less precise than most volume indicators, although still highly adequate for use as a quality indicator. Hospitals should examine more than one year of data if possible and average volumes for a more precise estimate. Hospitals may also consider use with the esophageal resection indicator, another complex cancer surgery. Most hospitals perform fewer than 10 procedures in a 5-year period; however, relatively strong relationships between volume and outcome—specifically post-operative mortality—have been noted in the literature.

Empirical evidence shows that a low percentage of procedures were performed at high-volume hospitals. At threshold 1, 30.3% of pancreatic resection procedures were performed at high-volume providers (and 5.1% of providers are high volume).[27] At threshold 2, 27.0% were performed at high-volume providers (and 4.2% of providers are high volume).[28 29]

[27]Glasgow RE, Mulvihill SJ. Hospital volume influences outcome in patients undergoing pancreatic resection for cancer. West J Med 1996;165(5):294-300.
[28]Glasgow, Mulvihill, 1996.
[29]Nationwide Inpatient Sample and State Inpatient Databases. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/data/hcup

Limitations on Use

As a volume indicator, pancreatic resection is a proxy measure for quality and should be used with other indicators.

Details

Face validity: Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?

The face validity of pancreatic resection depends on whether a strong association with outcomes of care is both plausible and widely accepted in the professional community. No recommendations regarding minimum procedure volume exist.

Precision: Is there a substantial amount of provider or community level variation that is not attributable to random variation?

Pancreatic resection is measured accurately with discharge data. Most facilities perform 10 or fewer pancreatectomies for cancer during a 5-year period; therefore, this indicator is expected to have poor precision.

Minimal bias: Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk *adjustment and statistical methods to remove most or all bias?*

Risk adjustment is not appropriate, because volume measures are not subject to bias due to disease severity and comorbidities.

Construct validity: Does the indicator perform well in identifying true (or actual) quality of care problems?

Higher volumes have been repeatedly associated with better outcomes after pancreatic surgery, although these findings may be limited by inadequate risk adjustment of the outcome measure.

One study used clinical data to estimate the association between hospital volume and mortality following pancreatic cancer surgery. Begg et al. analyzed retrospective data from the Surveillance, Epidemiology, and End Results (SEER)-Medicare linked database from 1984 through 1993.[30] The crude 30-day mortality rate was 12.9% at hospitals performing 1-5 pancreatic resections during the study period, versus 7.7% and 5.8% at hospitals performing 6-10 and 11 or more procedures, respectively. The association between volume and mortality remained highly significant (p<.001) in a multivariate model, adjusting for comorbidities, cancer stage and volume, and age.

Lieberman et al. used 1984-91 hospital discharge data from New York State to analyze the association between mortality after pancreatic cancer resection and hospital volumes.[31] Adjusting for the year of surgery, age, sex, race, payer source, transfer status, and the total number of secondary diagnoses, the standardized mortality rate was 19% at minimal-volume hospitals (fewer than 10 patients during the study period); 12% at low-volume hospitals (10-50 patients); 13% at medium-volume hospitals (51-80 patients); and 6% at high-volume hospitals (more than 80 patients). Studies using data from Ontario and Medicare data have generated similar results.[32] [33]

Empirical evidence shows that pancreatic resection volume—after adjusting for age, sex, and APR-DRG—is independently and negatively correlated with mortality for pancreatic resection (r=-.41, p<.001).[34]

Fosters true quality improvement: Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?

Low-volume providers may attempt to increase their volume without improving quality of care by performing the procedure on patients who may not qualify or benefit from the procedure. Additionally, shifting procedures to high-volume providers may impair access to care for certain types of patients.

Prior use: Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

Pancreatic cancer surgical volume has not been widely used as an indicator of quality.

---

[30] Begg CB, Cramer LD, Hoskins WJ, et al. Impact of hospital volume on operative mortality for major cancer surgery. JAMA 1998;280(20):1747-51.
[31] Lieberman MD, Kilburn H, Lindsey M, et al. Relation of perioperative deaths to hospital volume among patients undergoing pancreatic resection for malignancy. Ann Surg 1995;222(5):638-45.
[32] Simunovic M, To T, Theriault M, et al. Relation between hospital surgical volume and outcome for pancreatic resection for neoplasm in a publicly funded health care system [see comments]. Cmaj 1999;160(5):643-8.
[33] Birkmeyer JD, Finlayson SR, Tosteson AN, et al. Effect of hospital volume on in-hospital mortality with pancreaticoduodenectomy. Surgery 1999;125(3):250-6.
[34] Nationwide Inpatient Sample.

5.3 Abdominal Aortic Aneurysm Repair Volume (IQI 4)

Abdominal Aortic Aneurysm (AAA) repair is a relatively rare procedure that requires proficiency with the use of complex equipment; and technical errors may lead to clinically significant complications, such as arrhythmias, acute myocardial infarction, colonic ischemia, and death.

Relationship to Quality	Higher volumes have been associated with better outcomes, which represent better quality.
Benchmark	Threshold 1: 10 or more procedures per year Threshold 2: 32 or more procedures per year [37 38]
Definition	Raw volume of provider-level AAA repair.
Numerator	Discharges with ICD-9-CM codes of 3834, 3844, 3864 and 39.71in any procedure field with a diagnosis code of AAA in any field. Age 18 years and older. Exclude cases: • MDC 14 (pregnancy, childbirth, and puerperium) • MDC 15 (newborns and other neonates)
Denominator	Not applicable.
Type of Indicator	Provider Level, Procedure Volume Indicator
Empirical Rating	Not applicable.

Summary of Evidence

AAA repair volume is measured with great precision, although volume indicators overall are not direct measures of quality and are relatively insensitive. For this reason, this indicator should be used in conjunction with other measures of mortality to ensure that increasing volumes truly improve patient outcomes. The volume-outcome relationship on which this indicator is based may not hold over time, as providers become more experienced or as technology changes.

As noted in the literature, higher volume hospitals have lower mortality than lower volume hospitals, and the differences in patient case-mix do not account fully for these relationships.

Empirical evidence shows that a moderate to low percentage of procedures were performed at high-volume hospitals, depending on which threshold is used. At threshold 1, 83.9% of AAA repair procedures were performed at high-volume providers (and 44.3% of providers are high volume). At threshold 2, 43.0% were performed at high-volume providers (and 12.2% of providers are high volume).[35 36 37 38]

Limitations on Use

As a volume indicator, AAA repair is a proxy measure for quality and should be used with other indicators.

Details

Face validity: Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?

The face validity of AAA repair depends on whether a strong association with outcomes of care is widely accepted in the professional community. No consensus recommendations about minimum procedure volume currently exist.

Precision: Is there a substantial amount of provider or community level variation that is not attributable to random variation?

[35]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 1992;27(4):517-42.

[36]Kazmers A, Jacobs L, Perkins A, et al. Abdominal aortic aneurysm repair in Veterans Affairs medical centers. J Vasc Surg 1996;23(2):191-200.

[37]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.

[38]Nationwide Inpatient Sample and State Inpatient Databases. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD.http://www.ahrq.gov/data/hcup

AAA repair is an uncommon cardiovascular procedure—only 48,600 were performed in the United States in 1997.[39] Although AAA repair is measured accurately with discharge data, the relatively small number of procedures performed annually at most hospitals suggests that volume may be subject to much random variation.

Minimal bias: Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?

Risk adjustment is not appropriate, because volume measures are not subject to bias due to disease severity and comorbidities.

Construct validity: Does the indicator perform well in identifying true (or actual) quality of care problems?

Most studies published since 1985 showed a significant association between either hospital or surgeon volume and inpatient mortality after AAA repair, although these findings may be limited by inadequate risk adjustment of the outcome measure and differ by type of aneurysms (intact vs. ruptured) being considered.

Several studies have explored whether experience on related, but not identical, cases may lead to improved outcomes. One study found that hospital volume of surgery for ruptured aneurysms was not associated with postoperative inpatient mortality, but it was associated with fewer inpatient deaths for ruptured aneurysms, suggesting that high-volume hospitals may manage ruptured aneurysms more aggressively.[40] One study that evaluated the impact of total vascular surgery volume found a significant effect for both ruptured and intact aneurysms.[41]

Empirical evidence shows that AAA repair volume and mortality—after adjusting for age, sex, and APR-DRG—are independently and negatively correlated with each other ($r=-.35$, $p<.001$).[42]

Fosters true quality improvement: Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?

Low-volume providers may attempt to increase their volume without improving quality of care by performing the procedure on patients who may not qualify or benefit. Additionally, shifting procedures to high-volume providers may impair access to care for certain types of patients.

Prior use: Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

The Center for Medical Consumers posts volumes of "resection of aorta with replacement" for New York hospitals.[43] The Pacific Business Group on Health states that "one marker of how well a hospital is likely to perform is...the number of (AAA) surgeries a hospital performs."[44]

[39] HCUPnet. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/data/hcup/.

[40] Kantonen I, Lepantalo M, Brommels M, et al. Mortality in ruptured abdominal aortic aneurysms. The Finnvasc Study Group. . Eur J Vasc Endovasc Surg 1999;17(3):208-12.

[41] Amundsen S, Skjaerven R, Trippestad A, et al. 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 Chir Scand 1990;156(4):323-7; discussion 327-8.

[42] Nationwide Inpatient Sample.

[43] The Center for Medical Consumers. (http://www.medicalconsumers.org/)

[44] http://www.pbgh.org/

5.4 Coronary Artery Bypass Graft Volume (IQI 5)

Coronary artery bypass graft (CABG) requires proficiency with the use of complex equipment; and technical errors may lead to clinically significant complications, such as myocardial infarction, stroke, and death.

Relationship to Quality	Higher volumes have been associated with better outcomes, which represent better quality.
Benchmark	Threshold 1: 100 or more procedures per year Threshold 2: 200 or more procedures per year
Definition	Raw volume of provider-level CABG.
Numerator	Discharges with ICD-9-CM codes of 3610 through 3619 in any procedure field. Age 18 years and older. Exclude cases: • MDC 14 (pregnancy, childbirth, and puerperium) • MDC 15 (newborns and other neonates)
Denominator	Not applicable.
Type of Indicator	Provider Level, Procedure Volume Indicator
Empirical Rating	Not applicable.

Summary of Evidence

CABG is measured with great precision, although volume indicators overall are not direct measures of quality and are relatively insensitive. For this reason, CABG should be used in conjunction with other measures of mortality to ensure that increasing volumes truly improve patient outcomes.

As noted in the literature, higher volumes of CABG have been associated with fewer deaths. However, the American Heart Association (AHA) and the American College of Cardiology (ACC) recommend that since some low-volume hospitals have very good outcomes, other measures besides volume should be used to evaluate individual surgeon's performance.

Empirical evidence shows that a high percentage of procedures were performed at high-volume hospitals. At threshold 1, 98.3% of CABG procedures were performed at high-volume providers (and 88% of providers are high volume).[45] At threshold 2, 90.7% were performed at high-volume providers (and 68% of providers are high volume).[46 47]

Limitations on Use

As a volume indicator, CABG is a proxy measure for quality and should be used with other indicators.

Details

Face validity: Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?

The face validity of CABG depends on whether a strong association with outcomes of care is both plausible and widely accepted in the professional community. The AHA and ACC have argued for "careful outcome tracking" and supported "monitoring institutions and individuals who annually perform fewer than 100 cases," although the panel noted that "some

[45] Eagle KA, Guyton RA, Davidoff R, et al. ACC/AHA Guidelines for Coronary Artery Bypass Graft Surgery: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Revise the 1991 Guidelines for Coronary Artery Bypass Graft Surgery). American College of Cardiology/American Heart Association. J Am Coll Cardiol 1999;34(4):1262-347.

[46] Hannan EL, Kilburn H, Jr., Bernard H, et al. Coronary artery bypass surgery: the relationship between inhospital mortality rate and surgical volume after controlling for clinical risk factors. Med Care 1991;29(11):1094-107.

[47] Nationwide Inpatient Sample and State Inpatient Databases. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/data/hcup

institutions and practitioners maintain excellent outcomes despite relatively low volumes."[48]

Precision: Is there a substantial amount of provider or community level variation that is not attributable to random variation?

CABG is measured accurately with discharge data. The large number of procedures performed annually at most hospitals suggests that annual volume is not subject to considerable random variation. Hannan et al. reported year-to-year hospital volume correlations of 0.96-0.97 in New York.[49]

Minimal bias: Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?

Risk adjustment is not appropriate, because volume measures are not subject to bias due to disease severity and comorbidities.

Construct validity: Does the indicator perform well in identifying true (or actual) quality of care problems?

Higher volumes have been repeatedly associated with better outcomes of care, although these findings may be limited by inadequate risk adjustment of the outcome measure.

Hannan found that the adjusted relative risk of inpatient death at high-volume hospitals (more than 200 cases per year) in 1989-92 was 0.84, compared with low-volume hospitals.[50] However, only 3.3% of patients in that study underwent CABG at a low-volume hospital. Analyses using instrumental variables suggested that much of the volume effect may be due to "selective referral" of patients to high-quality centers.[51] [52]

Empirical evidence shows that CABG volume and mortality—after adjusting for age, sex, and APR-DRG—is independently and negatively correlated with mortality for CABG (r=-.29, p<.001).[53]

Fosters true quality improvement: Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?

Low-volume providers may attempt to increase their volume without improving quality of care by performing the procedure on patients who may not qualify or benefit from the procedure. Additionally, shifting procedures to high-volume providers may impair access to care for certain types of patients.

Prior use: Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

Specific CABG volume thresholds have been suggested as "standards" for the profession. The Pacific Business Group on Health states that "one marker of how well a hospital is likely to perform is...the number of (CABG) surgeries a hospital performs."[54]

---

[48]Eagle et al. 1999.
[49]Hannan EL, Kilburn H Jr., Racz M, et al. Improving the outcomes of coronary artery bypass surgery in New York state. JAMA 1994;271(10):761-6.
[50]Hannan et al. 1994.
[51]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.
[52]Luft HS, Hunt SS, Maerki SC. The volume-outcome relationship: practice-makes-perfect or selective-referral patterns? Health Serv Res 1987;22(2):157-82.
[53]Nationwide Inpatient Sample.
[54]http://www.pbgh.org/

5.5 Percutaneous Transluminal Coronary Angioplasty Volume (IQI 6)

Percutaneous transluminal coronary angioplasty (PTCA) is a relatively common procedure that requires proficiency with the use of complex equipment, and technical errors may lead to clinically significant complications. The definition for PTCA mortality rate (IQI 30) is also noted below. The QI software calculates mortality for PTCA, so that the volumes for this procedure can be examined in conjunction with mortality. However, the mortality measure should not be examined independently, because it did not meet the literature review and empirical evaluation criteria to stand alone as its own measure.

Relationship to Quality	Higher volumes have been associated with better outcomes, which represent better quality.
Benchmark	Threshold 1: 200 or more procedures per year Threshold 2: 400 or more procedures per year
Definition	Raw volume of PTCA.
Numerator	Discharges with ICD-9-CM codes 0066, 3601, 3602, 3605 in any procedure field. Age 18 years and older. Exclude cases: • MDC 14 (pregnancy, childbirth, and puerperium) • MDC 15 (newborns and other neonates)
Denominator	Not applicable.
Type of Indicator	Provider Level, Procedure Volume Indicator
Empirical Rating	Not applicable.

5.6 PTCA Mortality Rate (IQI 30)

Relationship to Quality	Better processes of care may reduce short-term mortality, which represents better quality.
Definition	Number of deaths per 100 PTCAs.
Numerator	Number of deaths with a code of PTCA in any procedure field.
Denominator	Discharges with ICD-9-CM codes 0066, 3601, 3602, 3605 in any procedure field. Age 40 years and older. Exclude cases: • missing discharge disposition (DISP=missing) • transferring to another short-term hospital (DISP=2) • MDC 14 (pregnancy, childbirth, and puerperium) • MDC 15 (newborns and other neonates)
Type of Indicator	Provider Level, Mortality Indicator – Recommended for use only with the corresponding volume indicator above.
Empirical Performance	Population Rate (2003): 1.30 per 100 discharges at risk
Empirical Rating	Not available.

Summary of Evidence

PTCA is measured with great precision, although volume indicators overall are not direct measures of quality and are relatively insensitive. For this reason, PTCA should be used in conjunction with measures of mortality and quality of care within cardiac care to ensure that increasing volumes truly improve patient outcomes. As noted in the literature, higher volumes of PTCA have been associated with fewer deaths and post-procedural coronary artery bypass grafts (CABG).

Empirical evidence shows that a moderate to high percentage of procedures were performed at high-volume hospitals. At threshold 1, 95.7% of PTCA procedures were performed at high-volume providers (and 69% of the providers are high volume).[55] At threshold 2, 77.0% were performed at high-volume providers (and 42% of providers are high volume).[56 57]

Limitations on Use

As a volume indicator, PTCA is a proxy measure for quality and should be used with other indicators.

Details

Face validity: Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?

The face validity of PTCA depends on whether a strong association with outcomes of care is both plausible and widely accepted in the professional community. The American Heart Association (AHA) and the American College of Cardiology (ACC) have stated that "a significant number of cases per institution—at least 200 PTCA procedures annually—is essential for the maintenance of quality and safe care."[58] Providers may wish to examine rates by surgeon with this indicator.

Precision: Is there a substantial amount of provider or community level variation that is not attributable to random variation?

PTCA is an increasingly common procedure (16.7 per 10,000 persons in 1997[59]) and is measured accurately with discharge data. The large number of procedures performed annually at most hospitals suggests that annual volume is not subject to considerable random variation.

Minimal bias: Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?

Risk adjustment is not appropriate, because volume measures are not subject to bias due to disease severity and comorbidities.

Construct validity: Does the indicator perform well in identifying true (or actual) quality of care problems?

Higher volumes have been repeatedly associated with better outcomes of care, although these findings may be limited by inadequate risk adjustment of the outcome measure.

Using hospital discharge data to adjust for age, gender, multilevel angioplasty, unstable angina, and six comorbidities, one study found that high-volume hospitals had significantly lower rates of same-stay coronary artery bypass surgery (CABG) and inpatient mortality than low-volume hospitals.[60] Better studies based on clinical data systems (adjusting for left ventricular function) have confirmed higher risk-adjusted mortality and CABG rates at low-volume hospitals relative to high-volume hospitals.[61]

Empirical evidence shows that PTCA volume is negatively related to several other post-procedural mortality rates: CABG ($r=-.21$, $p<.001$), craniotomy ($r=-.200$, $p<.0001$), and abdominal aortic aneurysm (AAA) repair ($r=-.45$, $p<.0001$).[62]

---

[55] Ryan TJ, Bauman WB, Kennedy JW, et al. Guidelines for percutaneous transluminal coronary angioplasty. .A report of the American Heart Association/American College of Cardiology Task Force on Assessment of Diagnostic and Therapeutic Cardiovascular Procedures (Committee on Percutaneous Transluminal Coronary Angioplasty). Circulation 1993;88(6):2987-3007.

[56] Hannan EL, Racz M, Ryan TJ, et al. .Coronary angioplasty volume-outcome relationships for hospitals and cardiologists. JAMA 1997;277(11):892-8.

[57] Nationwide Inpatient Sample and State Inpatient Databases. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/data/hcup

[58] Ryan et al., 1993.

[59] Kozak LJ, Lawrence L. National Hospital Discharge Survey: annual summary, 1997. Vital Health Stat 13 1999(144):i-iv, 1-46.

[60] Ritchie JL, Maynard C, Chapko MK, et al. 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.

[61] Hannan et al. 1997.

[62] Nationwide Inpatient Sample.

Fosters true quality improvement: Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?

Low-volume providers may attempt to increase their volume without improving quality of care by performing the procedure on patients who may not qualify or benefit from the procedure. Additionally, shifting procedures to high-volume providers may impair access to care for certain types of patients.

Prior use: Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

PTCA volume has not been widely used as an indicator of quality, although specific volume thresholds have been suggested as "standards" for the profession.[63]

[63]Hirshfeld JW, Jr., Ellis SG, Faxon DP. Recommendations for the assessment and maintenance of proficiency in coronary interventional procedures: Statement of the American College of Cardiology. J Am Coll Cardiol 1998;31(3):722-43.

5.7 Carotid Endarterectomy Volume (IQI 7)

Carotid endarterectomy (CEA) is a fairly common procedure that requires proficiency with the use of complex equipment; and technical errors may lead to clinically significant complications, such as abrupt carotid occlusion with or without stroke, myocardial infarction, and death. The definition for CEA mortality rate (IQI 31) is also noted below. The QI software calculates mortality for CEA, so that the volumes for this procedure can be examined in conjunction with mortality. However, the mortality measure should not be examined independently, because it did not meet the literature review and empirical evaluation criteria to stand alone as its own measure.

Relationship to Quality	Higher volumes have been associated with better outcomes, which represent better quality.
Benchmark	Threshold 1: 50 or more procedures per year Threshold 2: 101 or more procedures per year
Definition	Raw volume of provider-level CEA.
Numerator	Discharges with ICD-9-CM codes of 3812 in any procedure field. Age 18 years and older. Exclude cases: • MDC 14 (pregnancy, childbirth, and puerperium) • MDC 15 (newborns and other neonates).
Denominator	Not applicable.
Type of Indicator	Provider Level, Procedure Volume Indicator
Empirical Rating	Not applicable.

5.8 CEA Mortality Rate (IQI 31)

Relationship to Quality	Better processes of care may reduce short-term mortality, which represents better quality.
Definition	Number of deaths per 100 CEAs.
Numerator	Number of deaths with a code of CEA in any procedure field.
Denominator	Discharges with ICD-9-CM codes of 3812 in any procedure field. Age 18 years and older. Exclude cases: • missing discharge disposition (DISP=missing) • transferring to another short-term hospital (DISP=2) • MDC 14 (pregnancy, childbirth, and puerperium) • MDC 15 (newborns and other neonates)
Type of Indicator	Provider Level, Mortality Indicator – Recommended for use only with the corresponding volume indicator above.
Empirical Performance	Population Rate (2003): 0.62 per 100 discharges at risk
Empirical Rating	Not available.

Summary of Evidence

CEA is measured with great precision, although volume indicators overall are not direct measures of quality and are relatively insensitive. For this reason, CEA should be used with other measures of mortality to ensure that increasing volumes truly improve patient outcomes. As noted in the literature, higher volume hospitals have lower mortality and post-operative stroke rates than lower volume hospitals.

Empirical evidence shows that a moderate percentage of procedures were performed at high-volume hospitals.[64] At threshold 1, 77.8% of CEA procedures were performed at high-volume providers (and 37% of providers are high volume).[65] At threshold 2, 51.0% were performed at high-volume providers (and 17% of providers are high volume).[66 67]

Limitations on Use

As a volume indicator, CEA is a proxy measure for quality and should be used with other indicators.

Details

Face validity: Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?

The face validity of CEA depends on whether a strong association with outcomes of care is both plausible and widely accepted in the professional community. Recent guidelines focus on monitoring surgical outcomes rather than promoting volume standards.[68]

Precision: Is there a substantial amount of provider or community level variation that is not attributable to random variation?
CEA is measured accurately with discharge data. Approximately 144,000 CEAs were performed in the United States in 1997.[69] Many hospitals perform relatively few procedures, suggesting that the actual annual count of procedures may not be a reliable guide to the number of procedures performed on an ongoing basis. In one study of Medicare beneficiaries, approximately 50% of CEAs were performed in hospitals that performed 21 or fewer operations per year.[70]

Minimal bias: Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?

Risk adjustment is not appropriate, because volume measures are not subject to bias due to disease severity and comorbidities.

Construct validity: Does the indicator perform well in identifying true (or actual) quality of care problems?

Although higher volumes have repeatedly been associated with better outcomes after CEA, these findings may be limited by inadequate risk adjustment of the outcome measure. Cebul et al. found that undergoing surgery in a high-volume hospital was associated with a 71% reduction in the risk of stroke or death at 30 days, after adjusting for age, gender, indication for surgery, renal insufficiency, and two cardiovascular comorbidities.[71] In the study by Karp et al., the risk of severe stroke or death was 2.6 times higher at the lowest-volume hospitals than at the highest-volume hospitals.[72] Empirical evidence shows that CEA volume is negatively correlated with several other mortality indicators: coronary artery bypass graft (CABG) ($r=-.26$, $p<.0001$), abdominal aortic aneurysm (AAA) repair ($r=-.38$, $p<.0001$), and craniotomy ($r=-.18$, $p<.0001$).[73]

Fosters true quality improvement: Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?

Low-volume providers may attempt to increase their volume without improving quality of care by performing the procedure on patients who may

[64] Nationwide Inpatient Sample and State Inpatient Databases, Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/data/hcup.
[65] Manheim LM, Sohn MW, Feinglass J, et al. Hospital vascular surgery volume and procedure mortality rates in California, 1982-1994. J Vasc Surg 1998;28(1):45-46.
[66] Hannan EL, Popp AJ, Tranmer B, et al. Relationship between provider volume and mortality for carotid endarterectomies in New York state. Stroke 1998;29(11):2292-7.
[67] Dudley RA, Johansen KL, Brand R, et al. Selective referral to high-volume hospitals: estimating potentially avoidable deaths. JAMA 2000;283(9):1159-66.
[68] Biller J, Feinberg WM, Castaldo JE, et al. Guidelines for carotid endarterectomy: a statement of healthcare professionals from a Special Writing Group of the Stroke Council, American Heart Association. Circulation 1998;97(5):501-9.
[69] Owings, MF, Lawrence L. Detailed diagnoses and procedures, National Hospital Discharge Survey, 1997. Vital Health Stat 13 199(145):1-157.
[70] Cebul RD, Snow RJ, Pine R, et al. Indications, outcomes, and provider volumes for carotid endarterectomy. JAMA 1998;279(16):1282-7.
[71] Cebul et al. 1998.
[72] Karp, HR, Flanders WD, Shipp CC, et al. Carotid endarterectomy among Medicare beneficiaries: a statewide evaluation of appropriateness and outcome. Stroke 1998;29(1):46-52.
[73] Nationwide Inpatient Sample.

not qualify. Additionally, shifting procedures to high-volume providers may impair access to care for certain types of patients.

Prior use: Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

The Center for Medical Consumers posts CEA volumes for New York hospitals.[74] The Pacific Business Group on Health states that "one marker of how well a hospital is likely to perform is...the number of (CEA) surgeries a hospital performs."[75]

[74]The Center for Medical Consumers. (http://www.medicalconsumers.org./)
[75]http://www.pbgh.org/

5.9 Esophageal Resection Mortality Rate (IQI 8)

Esophageal cancer surgery is a rare procedure that requires technical proficiency; and errors in surgical technique or management may lead to clinically significant complications, such as sepsis, pneumonia, anastomotic breakdown, and death.

Relationship to Quality	Better processes of care may reduce mortality for esophageal resection, which represents better quality care.
Benchmark	State, regional, or peer group average.
Definition	Number of deaths per 100 patients with discharge procedure code of esophageal resection.
Numerator	Number of deaths (DISP=20) with a code of esophageal resection in any procedure field and a diagnosis code of esophageal cancer in any field.
Denominator	Discharges with ICD-9-CM codes of 4240 through 4242 in any procedure field and a diagnosis code of esophageal cancer in any field. Age 18 years and older. Exclude cases: • missing discharge disposition (DISP=missing) • transferring to another short-term hospital (DISP=2) • MDC 14 (pregnancy, childbirth, and puerperium) • MDC 15 (newborns and other neonates)
Type of Indicator	Provider Level, Mortality Indicator for Inpatient Procedures
Empirical Performance	Population Rate (2003): 8.54 per 100 population at risk
Empirical Rating	8

Summary of Evidence

Esophageal resection is a complex cancer surgery, and studies have noted that providers with higher volumes have lower mortality rates. This suggests that providers with higher volumes have some characteristics, either structurally or with regard to processes, that influence mortality.

This procedure is performed only by a select number of hospitals, which may compromise the precision of the indicator. Providers may wish to examine several consecutive years to potentially increase the precision of this indicator.

Limitations on Use

Risk adjustment for clinical factors is recommended because of the confounding bias for esophageal resection. In addition, little evidence exists supporting the construct validity of this indicator.

Details

Face validity: Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?

The primary evidence for esophageal resection mortality as an indicator arises from the volume-outcome literature. The causal relationship between hospital volume and mortality is unclear, and the differing processes that may lead to better outcomes have not been identified.

Precision: Is there a substantial amount of provider or community level variation that is not attributable to random variation?

Esophageal resection is a relatively uncommon procedure; Patti et al. noted that most hospitals perform 10 or fewer procedures during a 5-year period.[76] The precision of this indicator may be improved by using several years of data. Empirical evidence shows that this indicator is precise, with a raw provider level mean of 20.2% and a substantial standard deviation of 36.6%.[77]

Relative to other indicators, a smaller percentage of the variation occurs at the provider level, rather than the discharge level. The signal ratio (i.e., the proportion of the total variation across providers that is truly related to systematic differences in provider performance rather than random variation) is low, at 8.9%, indicating that most of the observed differences in provider performance very likely do not represent true differences.

Minimal bias: Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?

Although no studies specifically addressed the need for risk adjustment, most of the volume-outcome studies published have used some sort of risk adjustment. Most of these studies used administrative data for risk adjustment.

Construct validity: Does the indicator perform well in identifying true (or actual) quality of care problems?

There is no evidence for the construct validity of esophageal resection beyond the volume-outcome relationship. Two studies examined hospital volume as compared to in-hospital mortality rates. Patti et al. found decreasing mortality rates across five volume categories (17% for 1-5 procedures, 19% for 6-10 procedures, 10% for 11-20 procedures, 16% for 21-30 procedures, and 6% for more than 30 procedures).[78] Gordan et al. combined all complex gastrointestinal procedures, finding that low-volume hospitals (11-20 procedures per year) had an adjusted odds of death of 4.0 as compared to the one high-volume hospital.[79]

Fosters true quality improvement: Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?

No evidence exists on whether or not this indicator would stimulate true improvement in quality; however, it is possible that high-risk patients may be denied surgery.

Prior use: Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

Esophageal resection has not been widely used as a quality indicator.

[76] Patti MG, Corvera CU, Glasgow RE, et al. A hospital's annual rate of esophagectomy influences the operative mortality rate. J Gastrointest Surg 1998;2(2):186-92.
[77] Nationwide Inpatient Sample and State Inpatient Databases. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/data/hcup/
[78] Patti et al., 1998.
[79] Gordan 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.

5.10 Pancreatic Resection Mortality Rate (IQI 9)

Pancreatic resection is a rare procedure that requires technical proficiency; and errors in surgical technique or management may lead to clinically significant complications, such as sepsis, anastomotic breakdown, and death.

Relationship to Quality	Better processes of care may reduce mortality for pancreatic resection, which represents better quality care.
Benchmark	State, regional, or peer group average.
Definition	Number of deaths per 100 patients with discharge procedure code of pancreatic resection.
Numerator	Number of deaths (DISP=20) with a code of pancreatic resection in any procedure field and a diagnosis code of pancreatic cancer in any field.
Denominator	Discharges with ICD-9-CM codes of 526 or 527 in any procedure field and a diagnosis code of pancreatic cancer in any field. Age 18 years and older. Exclude cases: • missing discharge disposition (DISP=missing) • transferring to another short-term hospital (DISP=2) • MDC 14 (pregnancy, childbirth, and puerperium) • MDC 15 (newborns and other neonates)
Type of Indicator	Provider Level, Mortality Indicator for Inpatient Procedures
Empirical Performance	Population Rate (2003): 6.60 per 100 population at risk
Empirical Rating	5

Summary of Evidence

Pancreatic resection is a complex cancer surgery, and studies have noted that providers with higher volumes have lower mortality rates for the procedure than providers with lower volumes. This suggests that providers with higher volumes have some characteristics, either structurally or with regard to processes, that influence mortality.

This procedure is performed only by a select number of hospitals, which may compromise the precision of the indicator. Providers may wish to examine several consecutive years to potentially increase the precision of this indicator.

Limitations on Use

Risk adjustment for clinical factors is recommended because of the confounding bias for pancreatic resection. In addition, little evidence exists supporting the construct validity of this indicator.

Details

Face validity: Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?

The primary evidence for pancreatic resection mortality as an indicator arises from the volume-outcome literature. The causal relationship between hospital volume and mortality is unclear, and the differing processes that may lead to better outcomes have not been identified.

Precision: Is there a substantial amount of provider or community level variation that is not attributable to random variation?

Pancreatic resection is a relatively uncommon procedure; Glasgow et al. found that most hospitals in California perform 10 or fewer procedures during a 5-year period.[80] However,

[80]Glasgow RE, Mulvihill SJ. Hospital volume influences outcome in patients undergoing pancreatic resection for cancer. West J Med 1996;165(5):294-300.

the mortality rate is high, ranging from 4% to 13%.[81] The precision of this indicator may be improved by using several years of data. Empirical evidence shows that this indicator is moderately precise, with a raw provider level mean of 15.4% and a standard deviation of 31.3%.[82]

Relative to other indicators, a higher percentage of the variation occurs at the provider level, rather than the discharge level. The signal ratio (i.e., the proportion of the total variation across providers that is truly related to systematic differences in provider performance rather than random variation) is low, at 16.5%, indicating that some of the observed differences in provider performance very likely do not represent true differences.

Minimal bias: Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?

Although no studies specifically addressed the need for risk adjustment, most of the volume-outcome studies published have used administrative data for risk adjustment.

Construct validity: Does the indicator perform well in identifying true (or actual) quality of care problems?

There is no evidence for the construct validity of pancreatic resection beyond the volume-outcome relationship. Ten studies examined hospital volume as compared to in-hospital mortality rates. Glasgow and Mulvihill estimated the following risk-adjusted mortality rates across hospital volume categories during the 5-year study period: 14% for 1-5 procedures, 10% for 6-10 procedures, 9% for 11-20 procedures, 7% for 21-30 procedures, 8% for 31-50 procedures, and 4% for over 50 procedures.[83] Leiberman et al. found that surgeon volume was less significantly associated with mortality (6-13% across three volume categories).[84]

Fosters true quality improvement: Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?

No evidence exists on whether or not this indicator would stimulate true improvement in quality; however, it is possible that high-risk patients may be denied surgery.

Prior use: Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

Pancreatic resection has not been widely used as a quality indicator.

[81] Begg CB, Cramer LD, Hoskins WJ et al. Impact of hospital volume on operative mortality for major cancer surgery. JAMA 1998;280(20):1747-51.

[82] Nationwide Inpatient Sample and State Inpatient Databases. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/data/hcup/

[83] Glasgow RE, Mulvihill SJ. Hospital volume influences outcome in patients undergoing pancreatic resection for cancer. West J Med 1996;165(5):294-300.

[84] Lieberman MD, Kilburn H, Lindsey M, et al. Relation of perioperative deaths to hospital volume among patients undergoing pancreatic resection for malignancy. Ann Surg 1995;222(5):638-45.

5.11 Abdominal Aortic Aneurysm Repair Mortality Rate (IQI 11)

Abdominal aortic aneurysm (AAA) repair is a relatively rare procedure that requires proficiency with the use of complex equipment; and technical errors may lead to clinically significant complications, such as arrhythmias, acute myocardial infarction, colonic ischemia, and death.

Relationship to Quality	Better processes of care may reduce mortality for AAA repair, which represents better quality care.
Benchmark	State, regional, or peer group average.
Definition	Number of deaths per 100 discharges with procedure code of AAA repair.
Numerator	Number of deaths (DISP=20) with a code of AAA repair in any procedure field and a diagnosis of AAA in any field.
Denominator	Discharges with ICD-9-CM codes of 3834, 3844, 3864, 3971 in any procedure field and a diagnosis code of AAA in any field. Age 18 years and older. Exclude cases: • missing discharge disposition (DISP=missing) • transferring to another short-term hospital (DISP=2) • MDC 14 (pregnancy, childbirth, and puerperium) • MDC 15 (newborns and other neonates)
Type of Indicator	Provider Level, Mortality Indicator for Inpatient Procedures
Empirical Performance	Population Rate (2003): 7.27 per 100 discharges at risk
Empirical Rating	8

Summary of Evidence

AAA repair is a technically difficult procedure with a relatively high mortality rate. Higher volume hospitals have been noted to have lower mortality rates, which suggests that some differences in the processes of care between lower and higher volume hospitals result in better outcomes.

Empirical analyses of demographic risk adjustment noted some potential bias for this indicator. Additional medical chart review or analyses of laboratory data may be helpful in determining whether more detailed risk adjustment is necessary. This indicator should also be considered with length of stay and transfer rates to account for differing discharge practices among hospitals.

Limitations on Use

Risk adjustment for clinical factors is recommended because of the confounding bias for AAA repair mortality rate. In addition, little evidence exists supporting the construct validity of this indicator.

Details

Face validity: Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?

Studies have reported 40-55% in-hospital mortality after emergent repair of ruptured aneurysms.[85] [86] [87] These data suggest that improved quality of care could have a substantial impact on public health.

Precision: Is there a substantial amount of provider or community level variation that is not attributable to random variation?

The relatively small number of AAA resections performed by each hospital suggests that

[85] 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. J Vasc Surg 1998;28(3):413-20.
[86] Kazmers A, Jacobs L, Perkins A, et al. Abdominal aortic aneurysm repair in Veterans Affairs medical centers. J Vasc Surg 1996;23(2):191-200.
[87] Rutledge R, Oller DW, Meyer AA, et al. A statewide, population-based time-series analysis of the outcome of ruptured abdominal aortic aneurysm. Ann Surg 1996;223(5):492-502.

mortality rates at the hospital level are likely to be unreliable. Empirical evidence shows that his indicator is precise, with a raw provider level mean of 21.5% and a substantial standard deviation of 26.8%.[88]

Relative to other indicators, a higher percentage of the variation occurs at the provider level, rather than the discharge level. The signal ratio (i.e., the proportion of the total variation across providers that is truly related to systematic differences in provider performance rather than random variation) is low, at 30.7%, indicating that some of the observed differences in provider performance likely do not represent true differences.

Minimal bias: Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?

The known predictors of in-hospital mortality include whether the aneurysm is intact or ruptured, age, female gender, admission through an emergency room, various comorbidities such as renal failure and dysrhythmias, and Charlson's comorbidity index.[89 90 91] In the absence of studies explicitly comparing models with and without additional clinical elements, it is difficult to assess whether administrative data contain sufficient information to remove bias.

Construct validity: Does the indicator perform well in identifying true (or actual) quality of care problems?

The correlation between hospital or physician characteristics and in-hospital mortality in most studies supports the validity of in-hospital mortality as a measure of quality.[92 93] Finally, excessive blood loss, which is a potentially preventable complication of surgery, has been identified as the most important predictor of mortality after elective AAA repair.[94]

Empirical evidence shows that AAA repair mortality is positively related to other post-procedural mortality measures, such as craniotomy (r=.28, p<.0001) and coronary artery bypass graft (CABG) (r=.17, p<.01).[95]

Fosters true quality improvement: Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?

All in-hospital mortality measures may encourage earlier post-operative discharge, and thereby shift deaths to skilled nursing facilities or outpatient settings. Another potential response would be to avoid operating on high-risk patients.

Prior use: Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

The Pennsylvania Health Care Cost Containment Council includes AAA repair in the "Other major vessel operations except heart (DRG 100)" indicator. It is also used by HealthGrades.com.

[88]Nationwide Inpatient Sample and State Databases. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/data/hcup/

[89]Manheim LM, Sohn MW, Feinglass J, et al. Hospital vascular surgery volume and procedure mortality rates in California, 1982-1994. J Vasc Surg 1998;28(1):45-56.

[90]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 1992;27(4):517-42.

[91]Wen SW, Simunovic M, Williams JI, et al. Hospital volume, calendar age, and short term outcomes in patients undergoing repair of abdominal aortic aneurysm: the Ontario experience, 1988-92. J Epidemiol Community Health 1996;50(2):207-13.

[92]Pearce WH, Parker MA, Feinglass J, et al. The importance of surgeon volume and training in outcomes for vascular surgical procedures. J Vasc Surg 1999;29(5):768-76.

[93]Rutledge et al., 1996.

[94]Pilcher DB, Davis JH, Ashikaga T, et al. Treatment of abdominal aortic aneurysm in an entire state over 7½ years. Am J Surg 1980;139(4):487-94.

[95]Nationwide Inpatient Sample.

5.12 Coronary Artery Bypass Graft Mortality Rate (IQI 12)

Coronary artery bypass graft (CABG) is a relatively common procedure that requires proficiency with the use of complex equipment; and technical errors may lead to clinically significant complications such as myocardial infarction, stroke, and death.

Relationship to Quality	Better processes of care may reduce mortality for CABG, which represents better quality care.
Benchmark	State, regional, or peer group average.
Definition	Number of deaths per 100 discharges with procedure code of CABG.
Numerator	Number of deaths (DISP=20) with a code of CABG in any procedure field.
Denominator	Discharges with ICD-9-CM codes of 3610 through 3619 in any procedure field. Age 40 years and older. Exclude cases: • missing discharge disposition (DISP=missing) • transferring to another short-term hospital (DISP=2) • MDC 14 (pregnancy, childbirth, and puerperium) • MDC 15 (newborns and other neonates)
Type of Indicator	Provider Level, Mortality Indicator for Inpatient Procedures
Empirical Performance	Population Rate (2003): 3.39 per 100 discharges at risk
Empirical Rating	5

Summary of Evidence

CABG mortality is one of the most widely used and publicized post-procedural mortality indicators. Demographics, comorbidities, and clinical characteristics of severity of disease are important predictors of outcome that may vary systematically by provider. Chart review may help distinguish comorbidities from complications.

This indicator should be considered with length of stay and transfer rates to account for differing discharge practices among hospitals. The use of smoothed estimates to help avoid the erroneous labeling of outlier hospitals is recommended.

Limitations on Use

Some selection of the patient population may lead to bias; providers may perform more CABG procedures on less clinically complex patients with questionable indications. Risk adjustment for clinical factors, or at a minimum APR-DRGs, is recommended because of the confounding bias of this indicator. Finally, the evidence for the construct validity of this indicator is limited.

Details

Face validity: Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?

Post-CABG mortality rates have recently become the focus of State public reporting initiatives.[96] Studies suggest that these reports serve as the basis for discussions between physicians and patients about the risks of cardiac surgery.

Precision: Is there a substantial amount of provider or community level variation that is not attributable to random variation?

Without applying hierarchical statistical models to remove random noise, it is likely that hospitals will be identified as outliers as a result of patient variation and other factors beyond the hospital's control. Empirical evidence shows that this indicator is precise, with a raw provider level

[96] Localio AR, Hamory BH, Fisher AC, et al. The public release of hospital and physician mortality data in Pennsylvania. A case study. Med Care 199;35(3):272-286.

mean of 5.1% and a standard deviation of 6.2%.[97]

Relative to other indicators, a lower percentage of the variation occurs at the provider level, rather than the discharge level. The signal ratio (i.e., the proportion of the total variation across providers that is truly related to systematic differences in provider performance rather than random variation) is moderate, at 54.5%, indicating that some of the observed differences in provider performance likely do not represent true differences.

Minimal bias: Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?

Based on studies using large databases, cardiac function, coronary disease severity, and the urgency of surgery appear to be powerful predictors of mortality.[98] Some of these risk factors are not available from administrative data.

Construct validity: Does the indicator perform well in identifying true (or actual) quality of care problems?

Numerous studies have reported an association between hospital volume and mortality after CABG surgery. However, experienced surgeons and surgical teams should be able to improve post-operative mortality by reducing aortic cross-clamp time, which has been repeatedly associated with post-operative mortality after adjusting for a variety of patient characteristics.[99] It is unknown how performance of these processes of care would affect hospital-level mortality rates.

Empirical evidence shows that CABG mortality is positively related to bilateral catheterization and negatively related to laparoscopic cholecystectomy.[100]

Fosters true quality improvement: Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?

Public reporting of CABG mortality rates may cause providers to avoid high-risk patients. Sixty-three percent of cardiothoracic surgeons surveyed in Pennsylvania reported that they were "less willing" to operate on the most severely ill patients since mortality data were released.[101] However, one study using Medicare data shows no evidence that cardiac surgeons in New York, which also reports CABG mortality rates, avoided high-risk patients.[102] All in-hospital mortality measures may encourage earlier post-operative discharge, shifting deaths to skilled nursing facilities or outpatient settings and causing biased comparisons across hospitals with different mean lengths of stay.

Prior use: Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

CABG mortality is publicly reported by California, New Jersey, New York, and Pennsylvania. Recent users of CABG mortality as a quality indicator include the University Hospital Consortium, the Joint Commission on Accreditation of Healthcare Organizations' (JCAHO's) IMSystem, Greater New York Hospital Association, the Maryland Hospital Association (as part of the Maryland QI Project) and HealthGrades.com.

[97] Nationwide Inpatient Sample and State Databases. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/data/hcup/

[98] Higgins TL, Estafanous FG, Loop FD, et al. Stratification of morbidity and mortality outcome by preoperative risk factors in coronary artery bypass patients. A clinical severity score. JAMA 1992;267(17):2344-8.

[99] Ottino G, Bergerone S, Di Leo M, et al. Aortocoronary bypass results: a discriminant multivariate analysis of risk factors of operative mortality. J Cardiovasc Surg (Torino) 1990;31(1):20-5.

[100] Nationwide Inpatient Sample.

[101] Hannan EL, Siu AL, Kumar D, et al. Assessment of coronary artery bypass graft surgery performance in New York. Is there a bias against taking high-risk patients? Med Care 1997;35(1):49-56.

[102] Peterson ED, DeLong ER, Jollis JG, et al. Public reporting of surgical mortality: a survey of new York State cardiothoracic surgeons. Ann Thorac surg 1999;68(4):1195-200; discussion 12-1-2.

5.13 Craniotomy Mortality Rate (IQI 13)

Craniotomy for the treatment of subarachnoid hemorrhage or cerebral aneurysm entails substantially high post-operative mortality rates.

Relationship to Quality	Better processes of care may reduce mortality for craniotomy, which represents better quality care.
Benchmark	State, regional, or peer group average.
Definition	Number of deaths per 100 discharges with DRG code for craniotomy (DRG 001, 002, 528, 529, 530, and 543), with and without comorbidities and complications.
Numerator	Number of deaths (DISP=20) with DRG code for craniotomy (DRG 001, 002, 528, 529, 530, and 543), age 18 years and older, with and without comorbidities and complications.
Denominator	All discharges with DRG code for craniotomy (DRG 001, 002, 528, 529, 530, and 543), with and without comorbidities and complications. Age 18 years and older. Exclude cases: • with a principle diagnosis of head trauma • missing discharge disposition (DISP=missing) • transferring to another short-term hospital (DISP=2)
Type of Indicator	Provider Level, Mortality Indicator for Inpatient Procedures
Empirical Performance	Population Rate (2003): 7.10 per 100 discharges at risk
Empirical Rating	6

Summary of Evidence

Craniotomy is a complex procedure. Providers with high rates have better outcomes, although this may be an artifact of patient selection.

This indicator is measured with good precision and very high provider systematic variation. Empirical analyses showed substantial bias for this indicator, particularly for age, and providers should risk-adjust for age and comorbidities. Medical chart reviews or analyses of laboratory tests can also be used to examine other patient characteristics that increase case-mix complexity.

Limitations on Use

Risk adjustment for clinical factors, or at a minimum APR-DRGs, is recommended because of the confounding bias for craniotomy. In addition, little evidence exists supporting the construct validity of this indicator.

Details

Face validity: Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?

Craniotomy requires technical skill and the ability to identify the most appropriate cases. Post-operative mortality rates for craniotomy—together with measures of volume and utilization—will give a comprehensive perspective on provider performance for this condition.

Precision: Is there a substantial amount of provider or community level variation that is not attributable to random variation?

Most providers perform relatively high numbers of procedures; post-operative mortality rates are also relatively high, averaging nearly 14% for patients over age 65.[103]

[103]Taylor CL, Yuan A, Selman WR, et al. Mortality rates, hospital length of stay, and the cost of treating subarachnoid hemorrhage in older patients: institutional and geographical differences. J Neurosurg 1997;86(4):583-8.

Empirical evidence shows that this indicator is precise, with a raw provider level mean of 16.2% and a substantial standard deviation of 18.5%.[104]

Relative to other indicators, a higher percentage of the variation occurs at the provider level, rather than the discharge level. The signal ratio (i.e., the proportion of the total variation across providers that is truly related to systematic differences in provider performance rather than random variation) is low, at 28.9%, indicating that most of the observed differences in provider performance likely do not represent true differences.

Minimal bias: Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?

Studies have shown that patients undergoing treatment for subarachnoid hemorrhage had significantly higher post-craniotomy mortality rates by age group (from 3% for those 23-39 years old to 17% for those over 70 years old).[105][106]

Older patients generally present with more severe illness on admission, including lower levels of consciousness, worse grade, thicker subarachnoid clot, intraventricular hemorrhage, and hydrocephalus. Older patients also present with higher comorbidity rates, including diabetes; hypertension; and pulmonary, myocardial, and cerebrovascular disease.

Construct validity: Does the indicator perform well in identifying true (or actual) quality of care problems?

Providers performing more than 30 procedures per year have lower mortality than providers performing fewer than 30, although the volume-outcome relationship may be a product of patient selection.[107] In one study, patients who were referred to a large medical center for subarachnoid hemorrhage were less likely to have died early and had fewer severe indications, including lower clinical grade, rate of coma, diastolic blood pressure, and younger patient age.[108]

Craniotomy appears to be positively related to mortality associated with abdominal aortic aneurysm (AAA) repair (r=.28, p<.0001), coronary artery bypass graft (CABG) (r=.23, p<.0001), and stroke (r=.49, p<.0001).[109]

Fosters true quality improvement: Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?

All in-hospital mortality measures may encourage earlier post-operative discharge, and thereby shift deaths to skilled nursing facilities or outpatient settings. This phenomenon may also lead to biased comparisons among hospitals with different mean lengths of stay.

Prior use: Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

The University Hospital Consortium uses post-operative mortality for craniotomy, non-trauma related, as a quality measure.

[104]Nationwide Inpatient Sample and State Inpatient Databases. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/data/hcup/
[105]Stachniak JB, Layon AJ, Day AL, et al. Craniotomy for intracranial aneurysm and subarachnoid hemorrhage. Is course, cost, or outcome affected by age? Stroke 1996;27(2):276-81.
[106]Lanzino G, Kassell NF, Germanson TP, et al. Age and outcome after aneurysmal subarachnoid hemorrhage: why do older patients fare worse? J Neurosurg 1996;85(3):410-8.
[107]Soloman 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.
[108]Whisnant JP, Sacco SE, O'Fallon WM, et al. Referral bias in aneurysmal subarachnoid hemorrhage. J Neurosurg 1993;78(5):726-32.
[109]Nationwide Inpatient Sample.

5.14 Hip Replacement Mortality Rate (IQI 14)

Total hip arthroplasty (without hip fracture) is an elective procedure performed to improve function and relieve pain among patients with chronic osteoarthritis, rheumatoid arthritis, or other degenerative processes involving the hip joint.

Relationship to Quality	Better processes of care may reduce mortality for hip replacement, which represents better quality care.
Benchmark	State, regional, or peer group average.
Definition	Number of deaths per 100 patients with discharge procedure code of partial or full hip replacement.
Numerator	Number of deaths (DISP=20) with a code of partial or full hip replacement in any procedure field.
Denominator	All discharges with procedure code of partial or full hip replacement in any field. Age 18 years and older. Include only discharges with uncomplicated cases: diagnosis codes for osteoarthrosis of hip in any field. Exclude cases: • missing discharge disposition (DISP=missing) • transferring to another short-term hospital (DISP=2) • MDC 14 (pregnancy, childbirth, and puerperium) • MDC 15 (newborns and other neonates)
Type of Indicator	Provider Level, Mortality Indicator for Inpatient Procedures
Empirical Performance	Population Rate (2003): 0.29 per 100 discharges at risk
Empirical Rating	3

Summary of Evidence

Hip replacement is an elective surgery with relatively low mortality rates. However, the main recipients of hip replacement are elderly individuals with increased risk for complications and morbidity from surgery.

Although the low mortality rate is likely to affect the precision of this indicator, the precision is adequate for a quality indicator. Patient characteristics such as age and comorbidities may influence the mortality rate. Risk adjustment is highly recommended for this indicator, and providers may want to examine the case mix of their populations. This indicator should be considered with length of stay and transfer rates to account for differing discharge practices among hospitals.

Limitations on Use

Because hip replacement is an elective procedure, some selection of patient population may create bias. Risk adjustment for clinical factors, or at a minimum APR-DRGs, is recommended because of the confounding bias for hip replacement. In addition, little evidence exists supporting the construct validity of this indicator.

Details

Face validity: Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?

Mortality for hip replacement is very low, as it should be for a procedure that is designed to improve function rather than extend survival. However, elderly patients are at a significant risk of post-operative complications such as pneumonia, osteomyelitis, myocardial ischemia, and deep vein thrombosis. If not recognized and effectively treated, complications may lead to life-threatening problems.

Precision: Is there a substantial amount of provider or community level variation that is not attributable to random variation?

Primary total hip arthroplasty is one of the most frequent types of major orthopedic surgery; about 160,000 were performed in the United States in 1998.[110] The relatively small number of deaths following total hip arthroplasty suggests that mortality rates are likely to be unreliable at the hospital level. Empirical evidence shows that this indicator is adequately precise, with a raw provider level mean of 1.2% and a substantial standard deviation of 5.7%.[111]

Relative to other indicators, a high percentage of the variation occurs at the provider level, rather than the discharge level. The signal ratio (i.e., the proportion of the total variation across providers that is truly related to systematic differences in provider performance rather than random variation) is low, at 20.0%, indicating that some of the observed differences in provider performance very likely do not represent true differences.

Minimal bias: Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?

Hip replacement has the potential for selection bias caused by the decision to select surgery. The known predictors of in-hospital mortality include age, hip fracture, and the presence of any significant comorbidity.[112] [113]

Construct validity: Does the indicator perform well in identifying true (or actual) quality of care problems?

Using administrative data without any risk adjustment, Lavernia and Guzman found no association between hospital volume and mortality following total hip arthroplasty.[114] However, surgeons with fewer than 10 cases per year showed a significant increase in the death rate, and hospitals with fewer than 10 cases per year showed a significant increase in complications.

One observational study attributed a decrease in post-operative mortality (from 0.36% in 1981-85 to 0.10% in 1987-91) to changes in perioperative care, such as reduced intraoperative blood loss, more aggressive arterial and oximetric monitoring, and increased use of epidural instead of general anesthesia.[115]

Fosters true quality improvement: Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?

All in-hospital mortality measures may encourage earlier post-operative discharge, and thereby shift deaths to skilled nursing facilities or outpatient settings.

Prior use: Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

Hip replacement was included in the original HCUP QIs; it is also used by HealthGrades.com and the Greater New York Hospital Association.

[110] Popovic JR, Kozak LJ. National hospital discharge survey: annual summary, 1998 [In Process Citation]. Vital Health Stat 13 2000(148):1-194.
[111] Nationwide Inpatient Sample. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://hcup.ahrq.gov/HCUPnet.asp.
[112] Kreder HF, Williams JI, Jaglal S, et al. 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.
[113] Whittle J, et al. 1993.
[114] Lavernia CJ, Guzman JF. Relationship of surgical volume to short-term mortality, morbidity, and hospital charges in arthroplasty. J Arthroplasty 1995;10(2):133-40.
[115] Sharrock et al. 1995.

5.15 Acute Myocardial Infarction Mortality Rate (IQI 15)

Timely and effective treatments for acute myocardial infarction (AMI), which are essential for patient survival, include appropriate use of thrombolytic therapy and revascularization.

Relationship to Quality	Better processes of care may reduce mortality for AMI, which represents better quality.
Benchmark	State, regional, or peer group average.
Definition	Number of deaths per 100 discharges with a principal diagnosis code of AMI.
Numerator	Number of deaths (DISP=20) with a principal diagnosis code of AMI.
Denominator	All discharges with a principal diagnosis code of AMI. Age 18 years and older. Exclude cases: • missing discharge disposition (DISP=missing) • transferring to another short-term hospital (DISP=2)
Type of Indicator	Provider Level, Mortality Indicator for Inpatient Conditions
Empirical Performance	Population Rate (2003): 8.85 per 100 discharges at risk
Empirical Rating	5

5.16 Acute Myocardial Infarction Mortality Rate, Without Transfer Cases (IQI 32)

Relationship to Quality	Better processes of care may reduce mortality for AMI, which represents better quality.
Benchmark	State, regional, or peer group average.
Definition	Number of deaths per 100 discharges with a principal diagnosis code of AMI.
Numerator	Number of deaths (DISP=20) with a principal diagnosis code of AMI.
Denominator	All discharges with a principal diagnosis code of AMI. Age 18 years and older. Exclude cases: • missing discharge disposition (DISP=missing) • transferring to another short-term hospital (DISP=2) • missing admission source (ASOURCE=missing) • transferring from another short-term hospital (ASOURCE=2)
Type of Indicator	Provider Level, Mortality Indicator for Inpatient Conditions
Empirical Performance	Population Rate (2003): 9.75 per 100 discharges at risk
Empirical Rating	Not available

Summary of Evidence

Reductions in the mortality rate for AMI on both the patient level and the provider level have been related to better processes of care. AMI mortality rate is measured with adequate precision, although some of the observed variance may not actually reflect true differences in performance. Risk adjustment may be important—particularly for the extremes. Otherwise, some providers may be mislabeled as outliers.

Two methods of calculating AMI mortality are included in the AHRQ QIs. The second method (IQI 32) was added in Revision 3, and reflected the desire of users to have an alternative method of measuring AMI mortality that excluded patients transferred from another hospital. IQI 32 excludes incoming transfers, however, doing so results in the loss of transferred AMI patients from any quality measurement (since outgoing transfers are already excluded). Therefore, some users may wish to use the AMI Mortality Rate to ensure the inclusion of all AMI patients.

Limitations on Use

Thirty-day mortality may be significantly different than in-hospital mortality, leading to information bias. This indicator should be considered in conjunction with length-of-stay and transfer rates. Risk adjustment for clinical factors (or, at a minimum, APR-DRGs) is recommended.

Details

Face validity: Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?

AMI affects 1.5 million people each year, and approximately one-third die in the acute phase of the heart attack.[116] Studies that show processes of care linked to survival improvements have resulted in detailed practice guidelines covering all phases of AMI management.[117]

Precision: Is there a substantial amount of provider or community level variation that is not attributable to random variation?

The precision of AMI mortality rate estimates may be problematic for medium and small hospitals. Empirical evidence shows that this indicator is precise, with a raw provider level mean of 24.4% and a standard deviation of 16.1%.[118]

Relative to other indicators, a higher percentage of the variation occurs at the provider level rather than the discharge level. The signal ratio (i.e., the proportion of the total variation across providers that is truly related to systematic differences in provider performance rather than random variation) is moderate, at 42.8%, indicating that some of the observed differences in provider performance likely do not represent true differences.

Minimal bias: Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?

Numerous studies have established the importance of risk adjustment for AMI patients. The most important predictors of short-term AMI mortality have been shown to include age, previous AMI, tachycardia, pulmonary edema and other signs of congestive heart failure, hypotension and cardiogenic shock, anterior wall and Q-wave infarction, cardiac arrest, and serum creatinine or urea nitrogen.

Using different risk adjustment methods or data sources (administrative versus clinical data) affects which specific hospitals are identified as outliers.[119] [120]

Construct validity: Does the indicator perform well in identifying true (or actual) quality of care problems?

When Meehan et al. evaluated coding accuracy, severity of illness, and process-based quality of care in Connecticut hospitals, they found that the hospitals with the highest risk-adjusted mortality had significantly lower utilization of beneficial therapies.[121]

In the California Hospital Outcomes Project, hospitals with low risk-adjusted AMI mortality were more likely to give aspirin within 6 hours of arrival in the emergency room, perform cardiac catheterization and revascularization procedures within 24 hours, and give heparin to prevent thromboembolic complications.[122]

---

[116] American Heart Association. Heart Attack and Stroke Facts: 1996 Statistical Supplement. Dallas, TX: American Heart Association; 1996.

[117] 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(3):890-911.

[118] Nationwide Inpatient Sample and State Inpatient Databases. . Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/data/hcup

[119] Landon B, Iezzoni LI, Ash AS, et al. Judging hospitals by severity-adjusted mortality rates: the case of CABG surgery. Inquiry 1996;33(2):155-66.

[120] Second Report of the California Hospitals Outcomes Project, May 1996, Acute Myocardial Infarction. Sacramento, CA: Office of Statewide Health Planning and Development; 1996.

[121] Meehan TP, Hennen J, Radford MJ, et al. Process and outcome of care for acute myocardial infarction among Medicare beneficiaries in Connecticut: a quality improvement demonstration project. Ann Intern Med 1995;122(12):928-36.

[122] Second Report of the California Hospitals Outcomes Project, May 1996. Acute Myocardial Infarction.

Empirical evidence shows that AMI mortality is correlated with bilateral catheterization (r=-.16, p<.0001), mortality for congestive heart failure (CHF) (r=.46, p<.0001), pneumonia (r=.46, p<.0001), coronary artery bypass graft (CABG) (r=.50, p<.0001), stroke (r=.40, p<.0001), and gastrointestinal hemorrhage (r=.38, p<.0001).[123]

Fosters true quality improvement: Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?

The use of AMI mortality as an indicator is unlikely to impede access to needed care. However, a few patients who fail to respond to resuscitative efforts may not be admitted if there is pressure to reduce inpatient mortality.

Prior use: Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

AMI mortality has been widely used as a hospital quality indicator by State health departments and the Joint Commission for the Accreditation of Healthcare Organizations (JCAHO).

AMI mortality measured by IQI 32 is closely related to the JCAHO indicator for AMI mortality. Unlike the existing indicator for AMI mortality (IQI #15), it excludes patients transferring from another short-term hospital and patients missing admission source. This indicator is NOT risk adjusted in the same manner as the JCAHO indicator and does not exclude hospice patients as the JCAHO indicator (due to inability to identify hospice patients in data).

Sacramento, CA: Office of Statewide Health Planning and Development; 1996.
[123]Nationwide Inpatient Sample.

5.17 Congestive Heart Failure Mortality Rate (IQI 16)

Congestive heart failure (CHF) is a progressive, chronic disease with substantial short-term mortality, which varies from provider to provider.

Relationship to Quality	Better processes of care may reduce short-term mortality, which represents better quality.
Benchmark	State, regional, or peer group average.
Definition	Number of deaths per 100 discharges with principal diagnosis code of CHF.
Numerator	Number of deaths (DISP=20) with a principal diagnosis code of CHF.
Denominator	All discharges with a principal diagnosis code of CHF. Age 18 years and older. Exclude cases • missing discharge disposition (DISP=missing) • transferring to another short-term hospital (DISP=2) • MDC 14 (pregnancy, childbirth, and puerperium) • MDC 15 (newborns and other neonates).
Type of Indicator	Provider Level, Mortality Indicator for Inpatient Conditions
Empirical Performance	Population Rate (2003): 4.33 per 100 discharges at risk
Empirical Rating	6

Summary of Evidence

CHF is a relatively common admission, with a relatively high short-term mortality rate. Certain procedures have been shown to decrease short-term CHF mortality on a patient level, but the impact of these practices on decreasing provider-level mortality is unknown.

CHF mortality has not been studied extensively as an indicator; however, some risk models have been developed that demonstrate the importance of comorbidities and some clinical factors in predicting death. Risk adjustment may be important—particularly for the extremes. Otherwise, some providers may be mislabeled as outliers.

Limitations on Use

CHF care occurs in an outpatient setting, and selection bias may be a problem for this indicator. In addition, 30-day mortality may be significantly different than in-hospital mortality, leading to information bias. Risk adjustment for clinical factors (or at a minimum APR-DRGs) is recommended.

Details

Face validity: Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?

Approximately 2 million persons in the United States have heart failure each year.[124] These numbers will likely increase as the population ages. The literature suggests that hospitals have improved care for heart failure patients. In a study of 29,500 elderly patients in Oregon, the 3-day mortality decreased by 41% from 1991 to 1995.[125]

The accuracy of ICD-9-CM coding for heart failure has been questioned. Although the specificity of a principal diagnosis of heart failure is high, the sensitivity is low.[126] Face validity will be maximized by limiting analyses to patients with a principal diagnosis of heart failure.

[124]Smith, WM. Epidemiology of congestive heart failure. Am J Cardiol 1985;55(2):3A-8A.
[125]Ni H, Hershberger FE. Was the decreasing trend in hospital mortality from heart failure attributable to improved hospital care? The Oregon experience, 1991-1995. Am J Manag Care 1999;5(9):1105-15.
[126]Goff, DC, Jr., Pandey DK, Chan FA, et al. Congestive heart failure in the United States: is there more than meets the I(CD code)? The Corpus Christi Heart Project. Arch Intern Med 2000;160(2):197-202.

Precision: Is there a substantial amount of provider or community level variation that is not attributable to random variation?

Empirical evidence shows that this indicator is precise, with a raw provider level mean of 7.5% and an standard deviation of 9.5%.[127]

Relative to other indicators, a lower percentage of the variation occurs at the provider level rather than the discharge level. The signal ratio (i.e., the proportion of the total variation across providers that is truly related to systematic differences in provider performance rather than random variation) is moderate, at 53.5%, indicating that some of the observed differences in provider performance likely do not represent true differences.

Minimal bias: Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?

Mortality is greatly influenced by age, transfer, cerebrovascular disease, chronic obstructive pulmonary disease, hyponatremia, other hydro-electrolytic disturbance, metastatic disease, renal disease, ventricular arrhythmia, liver disease, malignancy, hypotension, and shock.[128] [129] [130]

Construct validity: Does the indicator perform well in identifying true (or actual) quality of care problems?

No studies specifically examined the construct validity of in-hospital mortality from heart failure. Although processes of care have been shown to decrease mortality on a patient level, the effect of these processes of care on provider-level mortality rates is unknown.

Empirical evidence shows that CHF mortality is positively related to other mortality indicators, such as pneumonia, gastrointestinal hemorrhage, and stroke.

Fosters true quality improvement: Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?

Risk-adjusted measures of mortality may lead to an increase in coding of comorbidities. All in-hospital mortality measures may encourage earlier post-operative discharge, and thereby shift deaths to skilled nursing facilities or outpatient settings. However, Rosenthal et al. found no evidence that hospitals with lower in-hospital standardized mortality had higher (or lower) early post-discharge mortality.[131]

Prior use: Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

CHF mortality has been widely used as a quality indicator. HealthGrades.com, the University Hospital Consortium, and the Greater New York Hospital Association have used this measure. The Maryland Hospital Association includes this measure in its Maryland QI Project Indicator set. Likewise, the Michigan Hospital Association includes CHF in an aggregated mortality measure.

[127] Nationwide Inpatient Sample and State Inpatient Databases. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/data/hcup

[128] Yusuf, et al. 1989.

[129] MacIntyre K, Capewell lS, Stewart S, et al. Evidence of improving prognosis in heart failure: trends in case fatality in 66,547 patients hospitalized between 1986 and 1995 [see comments]. Circulation 2000;102(10):1126-31.

[130] Psaty BM, Boineau R, Kuller LH, et al. The potential costs of upcoding for heart failure in the United States. Am J Cardiol 1999;84(1):108-9, A9.

[131] Rosenthal GE, Baker DW, Norris DG, et al. Relationships between in-hospital and 30-day standardized hospital mortality: implications for profiling hospitals. Health Serv Res 2000;34(7):1449-68.

5.18 Acute Stroke Mortality Rate (IQI 17)

Quality treatment for acute stroke must be timely and efficient to prevent potentially fatal brain tissue death, and patients may not present until after the fragile window of time has passed.

Relationship to Quality	Better processes of care may reduce short-term mortality, which represents better quality.
Benchmark	State, regional, or peer group average.
Definition	Number of deaths per 100 discharges with principal diagnosis code of stroke.
Numerator	Number of deaths (DISP=20) with a principal diagnosis code of stroke.
Denominator	All discharges with a principal diagnosis code of stroke. Age 18 years and older. Exclude cases: • missing discharge disposition (DISP=missing) • transferring to another short-term hospital (DISP=2) • MDC 14 (pregnancy, childbirth, and puerperium) • MDC 15 (newborns and other neonates)
Type of Indicator	Provider Level, Mortality Indicator for Inpatient Conditions
Empirical Performance	Population Rate (2003): 11.16 per 100 discharges at risk
Empirical Rating	10

Summary of Evidence

Quality treatment for stroke must be timely and efficient to prevent brain tissue death. Clinical factors of severity at presentation, including use of mechanical ventilation on the first day, may vary by hospital and influence mortality. Providers with high rates may wish to examine the case mix for these potentially complicating factors.

Further, hospitals with rehabilitation programs may have higher mortality rates. Providers may want to use acute stroke mortality in conjunction with length of stay for their hospitals and for surrounding areas. Many deaths occur out of the hospital, suggesting that linkage to death records for patients post-discharge may be a good addition to this indicator.

Limitations on Use

Some stroke care occurs in an outpatient setting, and selection bias may be a problem for this indicator. In addition, 30-day mortality may be somewhat different than in-hospital mortality, leading to information bias. Risk adjustment for clinical factors (or at a minimum APR-DRGs) is recommended. Coding appears suboptimal for acute stroke and may lead to bias.

Details

Face validity: Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?

Stroke remains the third leading cause of death in the United States.[132] However, hospital care has a relatively modest impact on patient survival, and most stroke deaths occur after the initial acute hospitalization.

Precision: Is there a substantial amount of provider or community level variation that is not attributable to random variation?

Because stroke severity has a large effect on acute mortality, hospital mortality rates may be subject to considerable random variation. According to the literature, only 10-15% of stroke patients die during hospitalization.[133] Empirical evidence shows that this indicator is

[132]Centers for Disease Control and Prevention. Deaths: Final Data for 2003. Available at http://www.cdc.gov/nchs/products/pubs/pubd/hestats/finaldeaths03/finaldeaths03.htm.

[133]Brown RD, Whisnant JP, Sicks JD, et al. Stroke incidence, prevalence, and survival: secular trends in Rochester, Minnesota, through 1989. Stroke 1996;27(3):373-80.

precise, with a raw provider level mean of 21.3% and a standard deviation of 13.7%.[134]

Relative to other indicators, a higher percentage of the variation occurs at the provider level, rather than the discharge level. The signal ratio (i.e., the proportion of the total variation across providers that is truly related to systematic differences in provider performance rather than random variation) is moderate, at 51.9%, indicating that some of the observed differences in provider performance likely do not represent true differences.

Minimal bias: Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?

Williams et al. pooled the results of four studies that showed significant inaccuracies in ICD-9-CM codes for identifying stroke patients.[135] However, there are no studies documenting cross-hospital variations in these coding practices.

More patients with transient ischemic attacks (TIAs) are likely to be admitted to some hospitals because of the increased interest in the care of acute stroke patients.[136] Therefore, hospitals with more liberal admitting policies may appear to have lower mortality rates.

Coma at presentation and a history of previous stroke substantially increase the mortality of patients admitted with stroke.[137] Patients with prior aspirin use tend to have better outcomes.[138]

Construct validity: Does the indicator perform well in identifying true (or actual) quality of care problems?

Thrombolytic therapy has been shown to be beneficial in acute stroke; however, the small percentage of patients who receive this treatment suggests that it is likely to have only a modest impact on hospital mortality.[139] Empirical evidence shows that stroke mortality is positively related to mortality indicators for pneumonia, gastrointestinal hemorrhage, and congestive heart failure.

Fosters true quality improvement: Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?

All in-hospital mortality measures may encourage earlier post-operative discharge, thereby shifting deaths to skilled nursing facilities or outpatient settings. This may lead to biased comparisons among hospitals with different mean lengths of stay. "Overcoding" TIAs as strokes may also decrease stroke mortality rates.

Prior use: Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

Stroke mortality indicators have been used by the HealthGrades.com, University Hospital Consortium, Maryland Hospital Association Quality Indicators Project, and the Greater New York Hospital Association.

[134] Nationwide Inpatient Sample and State Inpatient Databases. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/data/hcup/

[135] Williams GR, Jiang JG, Matchar DB, et al. Incidence and occurrence of total (first-ever and recurrent) stroke. Stroke 1999;30(12):2523-8.

[136] Feinberg WM. Guidelines for the management of transient ischemic attacks. Ad Hoc Committee on Guidelines for the Management of Transient Ischemic Attacks of the Stroke Council, American Heart Association, Heart Dis Stroke 1994;3(5):275-83.

[137] Samsa GP, Bian J, Lipscomb J, et al. Epidemiology of recurrent cerebral infarction: a Medicare claims-based comparison of first and recurrent strokes on 2-year survival and cost. Stroke 1999;30(2):338-49.

[138] Kalra L, Perez I, Smithard DG, et al. Does prior use of aspirin affect outcome in ischemic stroke? Am J Med 2000;108(3):205-9.

[139] Tissue plasminogen activator for acute ischemic stroke. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. N Engl J Med 1995;333(24):1581-7.

5.19 Gastrointestinal Hemorrhage Mortality Rate (IQI 18)

Gastrointestinal (GI) hemorrhage may lead to death when uncontrolled, and the ability to manage severely ill patients with comorbidities may influence the mortality rate.

Relationship to Quality	Better processes of care may reduce mortality for GI hemorrhage, which represents better quality.
Benchmark	State, regional, or peer group average.
Definition	Number of deaths per 100 discharges with principal diagnosis code of GI hemorrhage.
Numerator	Number of deaths (DISP=20) with a principal diagnosis code of gastrointestinal hemorrhage.
Denominator	All discharges with principal diagnosis code for gastrointestinal hemorrhage. Age 18 years and older. Exclude cases: • missing discharge disposition (DISP=missing) • transferring to another short-term hospital (DISP=2) • MDC 14 (pregnancy, childbirth, and puerperium) • MDC 15 (newborns and other neonates)
Type of Indicator	Provider Level, Mortality Indicator for Inpatient Conditions
Empirical Performance	Population Rate (2003): 3.00 per 100 discharges at risk
Empirical Rating	5

Summary of Evidence

GI hemorrhage itself is rarely the cause of death, and the extreme influence of comorbidities on the survival rate of patients with GI hemorrhage—as well as the influence of age and timing of onset (pre- or post-hospitalization)—raises questions about the potential bias of this indicator.

Providers should risk-adjust for comorbidities. In addition, providers with high rates may want to examine their case-mix for higher complexity of cases (e.g., patients over 60, more comorbidities).

Hospital practices differ, with some hospitals discharging patients earlier than others. For this reason, this indicator should be considered in conjunction with length of stay and transfer rates.

Limitations on Use

Limited evidence supports the construct validity of this indicator. Risk adjustment for clinical factors, or at a minimum APR-DRGs, is recommended because of the substantial confounding bias for this indicator.

Details

Face validity: Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?

Admission for GI hemorrhage is fairly common, and mortality rates vary greatly. Lower mortality has been associated with more use of treatments such as early endoscopy (within 24-48 hours of presentation). Mortality rates on large population-based databases have not changed since the 1940s, although the ages and comorbidities of patients have increased.[140]

Precision: Is there a substantial amount of provider or community level variation that is not attributable to random variation?

Rates of mortality in GI hemorrhage vary from 0% to 29%, with most studies reporting rates of 3.5% to 11%. Empirical evidence shows that this indicator is precise, with a raw provider

[140] Rockall TA, Logan RF, Devlin HB, et al. Variation in outcome after acute upper gastrointestinal haemorrhage. The National Audit of Acute Upper Gastrointestinal Haemorrhage. Lancet 1995;346(8971):346-50.

mean of 4.6% and a standard deviation of 5.7%.[141]

Relative to other indicators, a lower percentage of the variation occurs at the provider level, rather than the discharge level. The signal ratio (i.e., the proportion of the total variation across providers that is truly related to systematic differences in provider performance rather than random variation) is low, at 20.2%, indicating that some of the observed differences in provider performance do not represent true differences in provider performance.

Minimal bias: Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?

Mortality from GI hemorrhage is highly influenced by patient comorbidities, as well as the nature and severity of the bleed itself. One study noted that some endoscopic findings, hemodynamic characteristics, and comorbidities were highly predictive of life-threatening events.[142] Another study tested the effect of risk adjustment on hospital ranking for gastrointestinal hemorrhage mortality. Risk adjusting for age, shock, and comorbidity changed 30 hospitals' rankings by more than 10. Adding diagnosis, endoscopy findings, and rebleed status changed 32 hospital rankings by more than 10.[143]

Construct validity: Does the indicator perform well in identifying true (or actual) quality of care problems?

No studies explicitly evaluated the construct validity of GI hemorrhage. Although processes of care have been shown to decrease mortality on a patient level, the effect of these processes of care on provider-level mortality rates is unknown.

Empirical evidence shows that GI hemorrhage is positively related to mortality indicators such as pneumonia, stroke, and congestive heart failure.[144]

One meta-analysis showed a slight advantage for early endoscopy.[145] Another study found that endoscopy was not related to mortality in either the bivariate or multivariate analyses.[146]

Fosters true quality improvement: Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?

Risk-adjusted measures of mortality may lead to an increase in coding of comorbidities. All in-hospital mortality measures may encourage earlier post-operative discharge, and thereby shift deaths to skilled nursing facilities or outpatient settings. This phenomenon may also lead to biased comparisons among hospitals with different mean lengths of stay.

Prior use: Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

GI hemorrhage is currently used by the Cleveland Choice Health Quality Choice. The Maryland Hospital Association includes this measure in its Maryland QI Project Indicator set. Likewise, the Michigan Hospital Association includes GI hemorrhage in an aggregated mortality measure.

[141] Nationwide Inpatient Sample and State Inpatient Databases. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/data/hcup

[142] Hay JA, Lyubashevsky E, Elashoff J, et al. Upper gastrointestinal hemorrhage clinical guideline determining the optimal hospital length of stay. Am J Med 1996;100(3):313-22.

[143] Rockall et al., 1995.

[144] HCUPnet, Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality, Rockville, MD. http://hcup.ahrq.gov/HCUPnet.asp

[145] Cook DJ, Guyatt GH, Salena BJ, et al. Endoscopic therapy for acute nonvariceal upper gastrointestinal hemorrhage: a meta-analysis. Gastroenterology 1992;102(1):139-48.

[146] Cooper GS, Chak A, Way LE, et al. Early endoscopy in upper gastrointestinal hemorrhage: associations with recurrent bleeding, surgery, and length of hospital stay. Gastrointest Endosc 1999;49(2):145-52.

5.20 Hip Fracture Mortality Rate (IQI 19)

Hip fractures, which are a common cause of morbidity and functional decline among elderly persons, are associated with a significant increase in the subsequent risk of mortality.

Relationship to Quality	Better processes of care may reduce mortality for hip fracture, which represents better quality.
Benchmark	State, regional, or peer group average.
Definition	Number of deaths per 100 discharges with principal diagnosis code of hip fracture.
Numerator	Number of deaths (DISP=20) with a principal diagnosis code of hip fracture.
Denominator	All discharges with a principal diagnosis code for hip fracture. Age 18 years and older. Exclude cases: • missing discharge disposition (DISP=missing) • transferring to another short-term hospital (DISP=2) • MDC 14 (pregnancy, childbirth, and puerperium) • MDC 15 (newborns and other neonates)
Type of Indicator	Provider Level, Mortality Indicator for Inpatient Conditions
Empirical Performance	Population Rate (2003): 3.18 per 100 discharges at risk
Empirical Rating	10

Summary of Evidence

Complications of hip fracture and other comorbidities lead to a relatively high mortality rate, and evidence suggests that some of these complications are preventable. Hip fracture mortality rate is measured with good precision, although some of the observed variance does not reflect true differences in performance. About 89% of hip fracture patients are elderly.

Patient age, sex, comorbidities, fracture site, and functional status are all predictors of functional impairment and mortality. Administrative data may not contain sufficient information for these risk factors.

Limitations on Use

Thirty-day mortality may be somewhat different than in-hospital mortality, leading to information bias. Mortality rates should be considered in conjunction with length of stay and transfer rates. Risk adjustment for clinical factors (or at a minimum APR-DRGs) is recommended. Limited evidence exists for the construct validity of this indicator.

Details

Face validity: Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?

Hip fractures are associated with a significant increase in the subsequent risk of mortality, which persists for a minimum of 3 months among the oldest and most impaired individuals.[147 148] Elderly patients often have multiple comorbidities and pre-fracture functional impairments. As a result, they are at significant risk of postoperative complications, which—if not recognized and effectively treated—can lead to life-threatening problems.

Precision: Is there a substantial amount of provider or community level variation that is not attributable to random variation?

The largest published study of in-hospital mortality reported a rate of 4.9% in 1979-88, which suggests that mortality rates are likely to

[147]Forsen L, Sogaard AJ, Meyer HE, et al. Survival after hip fracture: short- and long-term excess mortality according to age and gender. Osteoporos Int 1999;10(1):73-8.
[148]Wolinsky FD, Fitzgerald JF, Stump TE. The effect of hip fracture on mortality, hospitalization, and functional status: a prospective study. Am J Public Health 1997;87(3):398-403.

be relatively reliable at the hospital level.[149] Empirical evidence shows that this indicator is precise, with a raw provider level mean of 14.4% and a standard deviation of 16.0%.[150]

Relative to other indicators, a higher percentage of the variation occurs at the provider level, rather than the discharge level. The signal ratio (i.e., the proportion of the total variation across providers that is truly related to systematic differences in provider performance rather than random variation) is moderate, at 54.3%, indicating that some of the observed differences in provider performance likely do not represent true differences.

Minimal bias: Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?

Demographic predictors of in-hospital or 30-day mortality include age, male sex, and prior residence in a nursing home. Fracture site may be a significant predictor for long-term outcomes. Comorbidity predictors include malnutrition; venous, digestive, and cardiovascular diseases; neoplasms, disorientation or delirium, chronic obstructive pulmonary disease, the number of chronic medical conditions, prior hospitalization within 1 month, and the American Society of Anesthesiology physical status score.

Empirical analyses confirm that this indicator has some potential bias, and risk adjustment with age and sex and APR-DRGs is highly recommended. Chart review may identify differences in functional status or other clinical factors not accounted for in discharge data.

Construct validity: Does the indicator perform well in identifying true (or actual) quality of care problems?

One study demonstrated that Medicare patients with poor “process of care” had similar risk-adjusted 30-day mortality rates as patients with good process of care.[151] Nevertheless, there is substantial evidence that at least two major causes of death among hip fracture patients are partially preventable: pulmonary emboli and acute myocardial infarction.[152] Very little evidence supports an association between hospital volume and mortality following hip fracture repair.

Empirical evidence shows that hip fracture repair mortality is positively related to pneumonia, stroke, gastrointestinal hemorrhage, and congestive heart failure mortality.[153]

Fosters true quality improvement: Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?

All in-hospital mortality measures may encourage earlier post-operative discharge. Thirty-day mortality for hip fracture is substantially higher than in-hospital mortality in the largest published studies, suggesting that a relatively modest decrease in mean length of stay could significantly decrease inpatient mortality. Another potential effect would be to avoid operating on high-risk patients, although this seems unlikely.

Prior use: Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

In-hospital mortality following hip fracture repair has not been widely used as a quality indicator, although it is included within a University Hospital Consortium indicator (mortality for DRG 209).

[149] Myers AH, Robinson EG, Van Natta ML, et al. Hip fractures among the elderly: factors associated with in-hospital mortality. Am J Epidemiol 1991;134(10):1128-37.
[150] Nationwide Inpatient Sample and State Inpatient Databases. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/data/hcup/
[151] Kahn KL, Rogers WH, Rubenstein LV, et al. Measuring quality of care with explicit process criteria before and after implementation of the DRG-based prospective payment system. JAMA 1990;264(15):1969-73.
[152] Perez JV, Warwick DJ, Case CP, et al. Death after proximal femoral fracture—an autopsy study. Injury 1995;26(4):237-40.
[153] Nationwide Inpatient Sample.

5.21 Pneumonia Mortality Rate (IQI 20)

Treatment with appropriate antibiotics may reduce mortality from pneumonia, which is a leading cause of death in the United States.

Relationship to Quality	Inappropriate treatment for pneumonia may increase mortality.
Benchmark	State, regional, or peer group average.
Definition	Mortality in discharges with principal diagnosis code of pneumonia.
Numerator	Number of deaths (DISP=20) with a principal diagnosis code of pneumonia.
Denominator	All discharges with principal diagnosis code of pneumonia. Age 18 years and older. Exclude cases: • missing discharge disposition (DISP=missing) • transferring to another short-term hospital (DISP=2) • MDC 14 (pregnancy, childbirth, and puerperium) • MDC 15 (newborns and other neonates)
Type of Indicator	Provider Level, Mortality Indicator for Inpatient Conditions
Empirical Performance	Population Rate (2003): 7.75 per 100 discharges at risk
Empirical Rating	7

Summary of Evidence

Pneumonia admissions are fairly common, and hospitals and physicians vary in admission practices. The high degree of patient heterogeneity suggests that providers may be mislabeled as poor quality without risk adjustment.

Providers with particularly high and low mortality rates should examine the case-mix of their patients for comorbidities, age, and clinical characteristics. Chart reviews may be helpful in determining whether differences truly arise from quality of care, or from patient-level differences in coding, comorbidities, or severity of disease. Providers may also wish to examine rates of outpatient care, because some patients are treated in outpatient settings.

Limitations on Use

Pneumonia care occurs in an outpatient setting, and selection bias may be a problem for this indicator. In addition, 30-day mortality may be somewhat different than in-hospital mortality, leading to information bias. Risk adjustment for clinical factors (or at a minimum APR-DRGs) is recommended.

Details

Face validity: Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?

Pneumonia is the sixth leading cause of death in the United States.[154] Patient characteristics are relatively important predictors of in-hospital mortality, although the performance of specific processes of care may also lead to better patient outcomes.

Precision Is there a substantial amount of provider or community level variation that is not attributable to random variation?

The high degree of heterogeneity among patients admitted for pneumonia suggests that the mortality indicator will be imprecise. However, empirical evidence shows that this indicator is precise, with a raw provider level mean of 13.8% and a standard deviation of 10.2%.[155]

[154] Hoyert DL, Kochanek KD, Murphy SL. Deaths: final data for 1997. Natl Vital Stat Rep 1999;47(19):1-104.
[155] Nationwide Inpatient Sample and State Inpatient Databases. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/data/hcup/

Relative to other indicators, a higher percentage of the variation occurs at the provider level rather than the discharge level. The signal ratio (i.e., the proportion of the total variation across providers that is truly related to systematic differences in provider performance rather than random variation) is moderate, at 62.9%, indicating that some of the observed differences in provider performance likely do not represent true differences.

Minimal bias: Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?

Comparison of hospital death rates with population death rates suggests that selection bias due to differing thresholds for admitting patients with pneumonia influences observed hospital mortality rates for pneumonia.[156] Population death rates from pneumonia (in particular, non-inpatient deaths) may be an important supplement to indicators based on hospital mortality. Some important predictors of pneumonia outcome are not reliably captured in administrative databases, including the microbial etiology, certain radiographic patterns, and pre-hospital functional status.[157] [158]

Construct validity: Does the indicator perform well in identifying true (or actual) quality of care problems?

A recent study reported an association between choice of antibiotics and 3-day mortality for patients hospitalized with pneumonia.[159] More basic than the choice of a particular antibiotic regimen is the timely administration of any antibiotic to the patient, which bears a plausible connection to improved outcomes.[160]

Empirical evidence shows that pneumonia mortality is positively related to stroke, gastrointestinal hemorrhage, and congestive heart failure.[161]

Fosters true quality improvement: Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?

All in-hospital mortality measures may encourage earlier post-operative discharge, and thereby shift deaths to skilled nursing facilities or outpatient settings. This phenomenon may also lead to biased comparisons among hospitals with different mean lengths of stay.

Prior use: Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

Pneumonia mortality is used as an indicator by the University Hospital Consortium, Greater New York Hospital Association, HealthGrades.com, Maryland Hospital Association, the Pennsylvania Health Care Cost Containment Council, and the California Hospital Outcomes Project.

---

[156] Markowitz JS, Pashko S, Gutterman EM, et al. Death rates among patients hospitalized with community-acquired pneumonia: a reexamination with data from three states. Am J Public Health 1996;86(8 Pt 1):1152-4.

[157] Fine MJ, Smith MA, Carson CA, et al. Prognosis and outcomes of patients with community-acquired pneumonia. A meta-analysis. JAMA 1996;275(2):134-41.

[158] Davis RB, Iezzoni LI, Phillips RS, et al. Predicting in-hospital mortality. The importance of functional status information. Med Care 1995;33(9):906-21.

[159] Gleason PP, Heehan TP, Fine JM, et al. Associations between initial antimicrobial therapy and medical outcomes for hospitalized elderly patients with pneumonia. Arch Intern Med 1999;159(21):2562-72.

[160] Meehan TP, Fine MJ, Krumholz HM, et al. Quality of care, process, and outcomes in elderly patients with pneumonia. JAMA 1997;278(23):2080-4.

[161] Nationwide Inpatient Sample.

5.22 Cesarean Delivery Rate (IQI 21)

Cesarean delivery is the most common operative procedure performed in the United States and is associated with higher costs than vaginal delivery. Despite a recent decrease in the rate of Cesarean deliveries, many organizations have aimed to monitor and reduce the rate.

Relationship to Quality	Cesarean delivery has been identified as an overused procedure. As such, lower rates represent better quality.
Benchmark	State, regional, or peer-group average.
Definition	Provider-level number of Cesarean deliveries per 100 deliveries.
Numerator	Number of Cesarean deliveries, identified by DRG, or by ICD-9-CM procedure codes if they are reported without a 7491 hysterotomy procedure.
Denominator	All deliveries. Exclude cases: • abnormal presentation, preterm, fetal death, multiple gestation diagnosis codes • breech procedure codes
Type of Indicator	Provider Level, Procedure Utilization Indicator
Empirical Performance	Population Rate (2003): 24.47 per 100 discharges at risk
Empirical Rating	17

5.23 Primary Cesarean Delivery Rate (IQI 33)

Relationship to Quality	Cesarean delivery has been identified as an overused procedure. As such, lower rates represent better quality.
Benchmark	State, regional, or peer-group average.
Definition	Provider-level number of Cesarean deliveries per 100 deliveries.
Numerator	Number of Cesarean deliveries, identified by DRG, or by ICD-9-CM procedure codes if they are reported without a 7491 hysterotomy procedure.
Denominator	All deliveries. Exclude cases: • abnormal presentation, preterm, fetal death, multiple gestation diagnosis codes • breech procedure codes • previous Cesarean delivery diagnosis in any diagnosis field.
Type of Indicator	Provider Level, Procedure Utilization Indicator
Empirical Performance	Population Rate (2003): 15.26 per 100 discharges at risk
Empirical Rating	Not evaluated

Summary of Evidence

The rate of Cesarean delivery in the United States increased from 5.5% in 1970 to a high of 24.7% in 1988 and decreased to 20.7% in 1996.[162] A review of the literature indicates that risk adjustment affects the outlier status and rankings of as many as 25% of hospitals. Given these results, providers may want to examine the clinical characteristics of their populations when interpreting the results of this indicator.

Clinical characteristics such as prior Cesarean, parity, breech presentation, placental or cord complications, sexually transmitted diseases (STDs), infections, and birth weight have been shown to explain substantial variation in Cesarean delivery rates. Information regarding some of these factors may be available by linking maternal discharge records to birth records. Providers may also wish to break down

[162] Menard MK. Cesarean delivery rates in the United States. The 1990s. Obstet Gynecol Clin North Am 1999;26(2):275-86.

this indicator into primary and repeat Cesarean delivery rates. Empirical analyses demonstrated that Cesarean delivery rate is measured with good precision.

Indicators for both total and primary cesarean delivery were included in Revision 3 of the AHRQ IQIs. Recently, the principle focus of quality initiatives has been primary cesarean deliveries, as more scrutiny has evolved around vaginal birth after cesarean delivery. However, some users, particularly when comparing with historical data, may wish to examine both the primary and total cesarean delivery rate.

Limitations on Use

Potential additional bias may result from clinical differences not identifiable in administrative data, so supplemental risk adjustment with linked birth records or other clinical data may be desirable. As a utilization indicator, the construct validity relies on the actual inappropriate use of procedures in hospitals with high rates, which should be investigated further.

Face validity: Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?

While the appropriateness of Cesarean delivery depends largely on patients' clinical characteristics, studies have shown that individual physician practice patterns account for a significant portion of the variation in Cesarean delivery rates.[163] [164] Non-clinical factors such as patient insurance status, hospital characteristics, and geographic region have also been related to rates.[165] [166] [167]

Precision: Is there a substantial amount of provider or community level variation that is not attributable to random variation?

Based on empirical evidence, this indicator is precise, with a raw provider level mean of 21.4% and a substantial standard deviation of 8.7%.[168]

Relative to other indicators, a higher percentage of the variation occurs at the provider level rather than the discharge level. However, the signal ratio (i.e., the proportion of the total variation across providers that is truly related to systematic differences in provider performance rather than random variation) is high, at 88.2%, indicating that the observed differences in provider performance represent true differences.

Minimal bias: Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?

The overall Cesarean delivery rate cannot determine appropriate use, but the variation in rates across institutions and regions may, if the variations do not merely reflect variations in patient disease severity and comorbidities.

Aron et al. used data from standardized reviews of medical records to adjust for clinical risk factors in women without prior Cesarean section. They found that hospital rankings often changed after risk adjustment, and in 57% of hospitals, the relative difference in unadjusted and adjusted rates was greater than 10%.[169] Additional studies found that risk-adjusting primary Cesarean delivery rates using a State birth certificate database substantially changes how hospital performance is judged.[170]

Construct validity: Does the indicator perform well in identifying true (or actual) quality of care problems?

The Cesarean rate for "optimal" quality of care is unknown, and many studies note that lower Cesarean rates do not necessarily reflect better quality care. Based on empirical evidence,

[163] Goyert GL, Bottoms FS, Treadwell MC, et al. The physician factor in cesarean birth rates [see comments]. N Engl J Med 1989;320(11):706-9.

[164] Berkowitz GS, Fiarman GS, Mojica MA, et al. Effect of physician characteristics on the cesarean birth rate [see comments]. Am J Obstet Gynecol 1989;161(1):146-9.

[165] Stafford RS. The impact of nonclinical factors on repeat cesarean section [see comments]. JAMA 1991;265(1):59-63.

[166] Haas JS, Udvarhelyi S, Epstein AM. The effect of health coverage for uninsured pregnant women on maternal health and the use of cesarean section [see comments]. JAMA 1993;270(1):61-4.

[167] Stafford RS, Sullivan SD, Gardner LB. Trends in cesarean section use in California, 1983 to 1990. Am J Obstet Gynecol 1993;168(4):1297-302.

[168] Nationwide Inpatient Sample and State Inpatient Databases. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/data/hcup

[169] Aron DC, Harper DL, Shepardson LB, et al. Impact of risk-adjusting cesarean delivery rates when reporting hospital performance. JAMA 1998;279(24):1968-72.

[170] Balit JL, Dooley SL, Peaceman AN. Risk adjustment for interhospital comparison of primary cesarean rates. Obstet Gynecol 1999;93(6):1025-30.

Cesarean delivery rate is inversely related to vaginal delivery after Cesarean (VBAC).[171]

Fosters true quality improvement: Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?

The Cesarean delivery rate can be decreased by decreasing the primary Cesarean delivery rate or increasing the VBAC rate. Sachs et al. note that when a trial of labor after Cesarean delivery fails, the rate of maternal morbidity, including infection and operative injuries, increases substantially.[172]

Prior use: Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

Cesarean delivery was included in the original HCUP QIs, and the reduction of Cesarean delivery rate is a goal for Healthy People 2010.[173]

Cesarean Delivery Rate (IQI #21) closely mirrors indicators used by Healthy People 2010 and American College of Obstetricians and Gynecology. Primary Cesarean Delivery Rate (IQI #33) mirrors the Joint Commission on the Accrediation of Healthcare Organizations (JCAHO) measure for Cesarean Delivery. Note that this indicator does not specifically exclude abortion procedures as the JCAHO measure does, although most abortion patients would not be included in the denominator.

---

[171] Nationwide Inpatient Sample.
[172] Sachs BP, Kobelin C, Castro MA, et al. The risks of lowering the cesarean-delivery rate. N Engl J Med 1999;340(1):54-7.
[173] Healthy People 2010. Office of Disease Prevention and Health Promotion, U.S. Department of Health and Human Services.

5.24 Vaginal Birth after Cesarean Rate, Uncomplicated (IQI 22)

The policy of recommending vaginal birth after Cesarean delivery (VBAC) represents to some degree a matter of opinion on the relative risks and benefits of a trial of labor in patients with previous Cesarean delivery.

Relationship to Quality	VBAC has been identified as a potentially underused procedure. As such, higher rates represent better quality.
Benchmark	State, regional, or peer-group average.
Definition	Provider-level vaginal births per 100 discharges with a diagnosis of previous Cesarean delivery.
Numerator	Number of vaginal births in women with a diagnosis of previous Cesarean delivery.
Denominator	All deliveries with a previous Cesarean delivery diagnosis in any diagnosis field. Exclude cases: • abnormal presentation, preterm, fetal death, multiple gestation diagnosis codes • breech procedure codes
Type of Indicator	Provider Level, Procedure Utilization Indicator
Empirical Performance	Population Rate (2003): 15.30 per 100 discharges at risk
Empirical Rating	19

5.25 Vaginal Birth after Cesarean Rate, All (IQI 34)

Relationship to Quality	VBAC has been identified as a potentially underused procedure. As such, higher rates represent better quality.
Benchmark	State, regional, or peer-group average.
Definition	Provider-level vaginal births per 100 discharges with a diagnosis of previous Cesarean delivery.
Numerator	Number of vaginal births in women with a diagnosis of previous Cesarean delivery.
Denominator	All deliveries with a previous Cesarean delivery diagnosis in any diagnosis field.
Type of Indicator	Provider Level, Procedure Utilization Indicator
Empirical Performance	Population Rate (2003): 14.84 per 100 discharges at risk
Empirical Rating	Not evaluated

Summary of Evidence

Health People 2010 established a goal of indirectly increasing VBAC rates by decreasing Cesarean deliveries in women with previous Cesarean deliveries to 63%.[174]

This indicator is measured with very good precision, and it is likely that the observed differences represent true differences in provider performance rather than random variation.

According to the literature, some clinical factors—such as previous classic Cesarean delivery—may contraindicate VBAC, and this indicator should be risk-adjusted for these factors. Because these clinical factors may not be available in administrative data, linkage to birth records may provide for better risk adjustment.

The best rate for VBAC has not been established. This indicator should be used in conjunction with area rates, national rates, and complication rates (maternal uterine rupture and length of stay, neonatal length of stay) to assess whether a rate is truly too high or too low.

[174] Healthy People 2010. Office of Disease Prevention and Health Promotion. U.S. Department of Health and Human Services.

Limitations on Use

Selection bias due to patient preferences and other factors may impact performance on this indicator. As noted earlier, supplemental adjustment with linked birth records or other clinical data may be desirable to address bias from clinical differences not identifiable in administrative data.

Details

Face validity: Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?

Despite the widespread use of VBAC rates as a quality indicator, a randomized trial comparing a trial of labor with elective repeat Cesarean delivery has yet to appear. In addition, approximately one-third of patients prefer to pursue repeat Cesarean delivery.[175] Many physicians appear to consider Cesarean delivery preferable to vaginal delivery, given the potential complications of the former.[176]

Precision: Is there a substantial amount of provider or community level variation that is not attributable to random variation?

Empirical evidence shows that this indicator is very precise, with a raw provider level mean of 33.6% and a substantial standard deviation of 14.8%.[177] Relative to other indicators, a higher percentage of the variation occurs at the provider level rather than the discharge level. The signal ratio (i.e., the proportion of the total variation across providers that is truly related to systematic differences in provider performance rather than random variation) is high, at 83.1%. This indicates that the observed differences in provider performance likely represent true differences, although some of the observed difference is due to patient characteristics.

[175]Roberts RG, Bell HS, Wall EM, et al. Trial of labor or repeated cesarean section. The woman's choice. Arch Fam Med 1997;6(2):120-5.
[176]Al-Mufti R, McCarthy A, Fisk NM. Obstetricians' personal choice and mode of delivery [letter] [see comments]. Lancet 1996;347(9000):544.
[177]Nationwide Inpatient Sample and State Inpatient Databases. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/data/hcup

Minimal bias: Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?

A study using birth certificates suggests that administrative data accurately distinguish the current mode of delivery (vaginal vs. Cesarean delivery), but less accurately identify VBAC and primary Cesarean delivery.[178] In addition, administrative data sources do not include the clinical factors required to identify appropriate candidates for trial of labor.[179] As a result, the denominator for VBAC rates calculated using administrative data will include women with an accepted medical indication for repeat Cesarean delivery, as well as patients who make an informed decision not to pursue a trial of labor.[180]

Construct validity: Does the indicator perform well in identifying true (or actual) quality of care problems?

The likelihood that a patient will undergo VBAC correlates with certain provider and institutional variables, suggesting that certain providers are more likely to adapt to changes in policy or technology. Based on empirical results, VBAC rates are inversely related to Cesarean delivery.[181]

Fosters true quality improvement: Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?

Promotion of VBAC as a quality indicator has led to successful increases in the VBAC rate in some cases, but not in others.[182] [183]

[178]Green DC, Moore JM, Adams MM, et al. Are we underestimating rates of vaginal birth after previous cesarean birth? The validity of delivery methods from birth certificates. Am J Epidemiol 1998;147(6):581-6.
[179]Aron DC, Harper DL, Shepardson LB, et al. Impact of risk-adjusting cesarean delivery rates when reporting hospital performance. JAMA 1998;279(24):1968-72.
[180]Roberts RG, Bell HS, Wall EM, et al. Trial of labor or repeated cesarean section. The woman's choice. Arch Fam Med 1997;6(2):120-5.
[181]Nationwide Inpatient Sample.
[182]Kazandjian VA, Lied TR. Cesarean section rates: effects of participation in a performance measurement project. Jt Comm J Qual Improv 1998;24(4):187-96.

Prior use: Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

VBAC was included in the original HCUP QI indicator set. In addition, the Joint Commission on Accreditation of Healthcare Organizations (JCAHO) has selected VBAC as one of its core measures.

[183]Bickell NA, Zdeb MS, Applegate MS, et al. Effect of external peer review on cesarean delivery rates: a statewide program. Obstet Gynecol 1996;87(5 Pt 1):664-7.

5.26 Laparoscopic Cholecystectomy Rate (IQI 23)

Surgical removal of the gall bladder (cholecystectomy) performed with a laparoscope has been identified as an underused procedure. Laparoscopic cholecystectomy is associated with less morbidity in less severe cases.

Relationship to Quality	Laparoscopic cholecystectomy is a new technology with lower risks than open cholecystectomy (removal of the gall bladder). Higher rates represent better quality.
Benchmark	State, regional, or peer-group average.
Definition	Number of laparoscopic cholecystectomies per 100 cholecystectomies.
Numerator	Number of laparoscopic cholecystectomies (any procedure field).
Denominator	All discharges with any procedure code of cholecystectomy in any procedure field. Age 18 years and older. Include only discharges with uncomplicated cases: cholecystitis or cholelithiasis in any diagnosis field. Exclude cases: • MDC 14 (pregnancy, childbirth, and puerperium) • MDC 15 (newborns and other neonates)
Type of Indicator	Provider Level, Procedure Utilization Indicator
Empirical Performance	Population Rate (2003): 75.55 per 100 discharges at risk
Empirical Rating	20

Summary of Evidence

Cholecystectomy—surgical removal of the gall bladder—is now performed with a laparoscope in about 75% of uncomplicated cases.[184] This indicator has a high percentage of variation attributable to providers. According to the literature, laparoscopic cholecystectomy may need to be adjusted for clinical severity, age, and other factors, because the procedure may be contraindicated for some patients, and others may not be clinically severe enough to qualify for cholecystectomy at all. Too many procedures in patients without appropriate clinical indications may artificially inflate the laparoscopic cholecystectomy rate without improving quality.

Limitations on Use

Up to one-half or more of all cholecystectomies are performed on an outpatient basis, and providers should incorporate outpatient data if possible when interpreting this indicator. Additional bias may result from clinical differences not identifiable in administrative data, so supplemental risk adjustment using other clinical data may be desirable. As a utilization indicator, the construct validity relies on the actual appropriate use of procedures in hospitals with high rates, which should be investigated further.

Details

Face validity: Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?

Laparoscopic cholecystectomy is associated with less postoperative pain, lower patient-controlled morphine consumption, better postoperative pulmonary function and oxygen saturation, and quicker return to limited activity.[185 186]

[184] Southern Surgeons Club. A prospective analysis of 1518 laparoscopic cholecystectomies. NEJM 1991;324:1073-1078.

[185] McMahon AJ, Russell IT, Baxter JN, et al. Laparoscopic and minilaparotomy cholecystectomy: a randomised trial [see comment]. Lancet 1994;343(8890):135-8.

Laparoscopic cholecystectomy requires more technical skill than the open approach. Therefore, a higher rate for this procedure (as a proportion of all cholecystectomies) suggests that a hospital can rapidly achieve proficiency in up-to-date treatment methods.

Precision: Is there a substantial amount of provider or community level variation that is not attributable to random variation?

According to the literature, cholecystectomies are relatively common, so moderately precise estimates of differences in laparoscopic use can be obtained. Based on empirical evidence, this indicator is very precise, with a raw provider level mean of 66.2% and a substantial standard deviation of 19.2%.[187]

Relative to other indicators, a higher percentage of the variation occurs at the provider level, rather than the discharge level. The signal ratio (i.e., the proportion of the total variation across providers that is truly related to systematic differences in provider performance rather than random variation) is high, at 89.1%, indicating that the observed differences in provider performance likely represent true differences.

Minimal bias: Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?

As surgeons become more experienced in laparoscopic cholecystectomies, they are likely to perform the procedure on more difficult patients. In addition, higher risks of complications are associated with older age and the presence of common bile duct stones.[188] Patient referral patterns and other selection factors may lead to substantial differences in laparoscopy rates (as a proportion of all cholecystectomies) across hospitals. Empirical results show that age and sex adjustment does seem to disproportionately impact hospitals in the low extreme relative to those in the high extreme.

Use of inpatient data could be substantially biasing, in that it eliminates those cholecystectomies performed on an outpatient basis, most of which are likely to be laparoscopic.

Construct validity: Does the indicator perform well in identifying true (or actual) quality of care problems?

According to the literature, there is no evidence that hospitals that use the laparoscopic approach more frequently provide better quality of care, based on other measures.

Fosters true quality improvement: Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?

One concern with this indicator is that the advent of laparoscopic surgery has led to a substantial increase in the overall cholecystectomy rate, especially involving uncomplicated and elective patients.[189] Another concern is that the "optimal" rate for this procedure has not been defined, and incentives to increase use may have negative consequences if local physicians lack appropriate training and expertise.

Prior use: Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

Laparoscopic cholecystectomy was included in the original HCUP QI indicator set.

---

[186] McMahon AF, Russell IT, Ramsay G, et al. Laparoscopic and minilaparotomy cholecystectomy: a randomized trial comparing postoperative pain and pulmonary function. Surgery 1994;115(5):533-9.

[187] Nationwide Inpatient Sample and State Inpatient Databases. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/data/hcup

[188] Jatzko GR, Lisborg PH, Pertl AM, et al. Multivariate comparison of complications after laparoscopic cholecystectomy and open cholecystectomy. Ann Surg 1995;221(4):381-6.

[189] Escarce JJ, Chen W, Schwartz JS. Falling cholecystectomy thresholds since the introduction of laparoscopic cholecystectomy. JAMA 1995;273(20):1581-5.

5.27 Incidental Appendectomy in the Elderly Rate (IQI 24)

Removal of the appendix incidental to other abdominal surgery—such as urological, gynecological, or gastrointestinal surgeries—is intended to eliminate the risk of future appendicitis and to simplify any future differential diagnoses of abdominal pain.

Relationship to Quality	Incidental appendectomy among the elderly is contraindicated. As such, lower rates represent better quality.
Benchmark	State, regional, or peer-group average.
Definition	Number of incidental appendectomies per 100 elderly with intra-abdominal procedure.
Numerator	Number of incidental appendectomies (any procedure field).
Denominator	All discharges age 65 years and older with intra-abdominal procedure (based on DRGs). Exclude cases: • MDC 14 (pregnancy, childbirth, and puerperium) • MDC 15 (newborns and other neonates).
Type of Indicator	Provider Level, Procedure Utilization Indicator
Empirical Performance	Population Rate (2003): 2.30 per 100 discharges at risk
Empirical Rating	13

Summary of Evidence

Incidental appendectomy is contraindicated in the elderly population, because this population has both a lower risk for developing appendicitis and a higher risk of postoperative complications. Given the low rate of incidental appendectomies, the precision for this indicator may be lower than other indicators.

Empirical analyses found that this indicator is moderately precisely measured, and the bias with respect to provider differences is not likely to be high.

Limitations on Use

As a utilization indicator, the construct validity relies on the actual inappropriate use of procedures in hospitals with high rates, which should be investigated further.

Details

Face validity: Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?

For the population as a whole, evidence remains unclear whether the removal of the appendix increases risk of morbidity and mortality significantly, or whether it is worth any amount of extra risk, given the low risk for future appendicitis and the ease of treatment.

Andrew and Roty showed that incidental appendectomy was associated with a higher risk of wound infection (5.9% versus 0.9%) among cholecystectomy patients who were at least 50 years of age, but not among younger patients.[190] Based on this finding and the findings of Warren and colleagues, the risk of incidental appendectomy is believed to outweigh the benefits for elderly patients.[191] [192] [193] [194] [195]

Precision: Is there a substantial amount of provider or community level variation that is not attributable to random variation?

Fewer than one-third of surgery departments routinely perform incidental appendectomies,

[190] Andrew MH, Roty AR, Jr. Incidental appendectomy with cholecystectomy: is the increased risk justified? Am Surg 1987;53(10):553-7.

[191] Warren JL, Penberthy LT, Addiss DG, et al. Appendectomy incidental to cholecystectomy among elderly Medicare beneficiaries. Surg Gynecol Obstet 1993;177(3):288-94.

[192] Fisher KS, Ross DS. Guidelines for therapeutic decision in incidental appendectomy. Surg Gynecol Obstet 1990;171(1):95-8.

[193] Synder TE, Selanders JR. Incidental appendectomy—yes or no? A retrospective case study and review of the literature. Infect Dis Obstet Gynecol 1998;6(1)30-7.

[194] Wolff BG. Current status of incidental surgery. Dis Colon Rectum 1995;38(4):435-41.

[195] Nockerts SR, Detmer DE, Fryback, DG. Incidental appendectomy in the elderly? No. Surgery 1980;88(2):301-6.

and rates may be difficult to estimate with precision at the majority of hospitals where it is not a routine procedure.[196]

Based on empirical evidence, this indicator is precise, with a raw provider level mean of 2.7% and a standard deviation of 3.5%.[197] Relative to other indicators, a higher percentage of the variation occurs at the discharge level than for some indicators. The signal ratio (i.e., the proportion of the total variation across providers that is truly related to systematic differences in provider performance rather than random variation) is moderate, at 55.4%, indicating that some of the observed differences in provider performance do not represent true differences.

Minimal bias: Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?

Incidental appendectomy appears to be contraindicated in an elderly population; therefore, very few (if any) cases would be justified by patients' preoperative characteristics. Empirical evidence shows that this indicator performs well to very well on multiple measures of minimum bias, and risk adjustment does not appear to impact the extremes of the distribution substantially.

Construct validity: Does the indicator perform well in identifying true (or actual) quality of care problems?

Most of the available evidence appears to contraindicate incidental appendectomy in the elderly, and performance of the procedure is subject to patient and surgeon preference. Therefore, incidental appendectomy rates may correlate poorly with other measures of hospital performance.

Fosters true quality improvement: Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?

Incidental appendectomy does not generally affect hospital payment; therefore, widespread use of this indicator may lead to less frequent coding of the procedure when it is performed. A reduction in the rate of incidental appendectomy may lead to a subsequent increase in the incidence of acute appendicitis, although this risk is expected to be small for the elderly population.

Prior use: Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

Incidental appendectomy in the elderly is a provider-level utilization indicator in the original HCUP QI set.

[196] Neulander EZ, Hawke CK, Soloway MS. Incidental appendectomy during radical cystectomy: an interdepartmental survey and review of the literature. Urology 2000;56(2):241-4.

[197] Nationwide Inpatient Sample and State Inpatient Databases. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/data/hcup

5.28 Bilateral Cardiac Catheterization Rate (IQI 25)

Right-side coronary catheterization incidental to left-side catheterization has little additional benefit for patients without clinical indications for right-side catheterization.

Relationship to Quality	Bilateral catheterization is contraindicated in most patients without proper indications. As such, lower rates represent better quality.
Benchmark	State, regional, or peer-group average.
Definition	Provider level bilateral cardiac catheterizations per 100 discharges with procedure code of heart catheterization.
Numerator	Number of simultaneous right and left heart catheterizations (in any procedure field). Age 18 years and older. Include only coronary artery disease. Exclude cases: • with valid indications for right-sided catheterization in any diagnosis field
Denominator	All discharges with heart catheterization in any procedure field. Include only coronary artery disease. Exclude cases: • MDC 14 (pregnancy, childbirth, and puerperium) • MDC 15 (newborns and other neonates)
Type of Indicator	Provider Level, Procedure Utilization Indicator
Empirical Performance	Population Rate (2003): 7.13 per 100 discharges at risk
Empirical Rating	25

Summary of Evidence

Bilateral cardiac catheterization received one of the highest precision ratings. Provider level variation accounts for a relatively large portion of the total variation compared to other indicators, meaning that variation for this indicator is influenced less by discharge level variation (patient level) than total variation for other indicators. It is likely that the observed differences in provider performance represent true differences, rather than random variation.

Analyses of minimum bias identified very little bias in this indicator when adjusting for APR-DRGs.

Limitations on Use

Outpatient procedures may result in selection bias for this indicator and should be examined. In addition, as a utilization indicator, the construct validity relies on the actual inappropriate use of procedures in hospitals with high rates, which should be investigated further.

Details

Face validity: Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?

Left-sided catheterization provides very useful information about coronary anatomy, as well as left ventricular function and valvular anatomy. However, the clinical yield for right-sided catheterization, which is often performed at the same time, is extremely low. The American College of Cardiology (ACC) and the American Heart Association (AHA) published guidelines for cardiac catheterization laboratories stating that "without specific indications, routine right heart catheterizations...are unnecessary."[198]

[198]Pepine CJ, Allen HD, Bashore TM, et al. ACC/AHA guidelines for cardiac catheterization and cardiac catheterization laboratories. American College of

Precision: Is there a substantial amount of provider or community level variation that is not attributable to random variation?

This measure should be estimable with reasonable precision, given that more than 1.2 million inpatient cardiac catheterizations were performed in the United States in 1998.[199] Based on empirical evidence, this indicator is very precise, with a raw provider level mean of 19.3% and a substantial standard deviation of 20.0%.[200]

Relative to other indicators, a higher percentage of the variation occurs at the provider level, rather than the discharge level. The signal ratio (i.e., the proportion of the total variation across providers that is truly related to systematic differences in provider performance rather than random variation) is very high, at 96.2%, indicating that the observed differences in provider performance likely represent true differences.

Minimal bias: Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?

Bilateral cardiac catheterization is considered appropriate in the presence of certain clinical indications: suspected pulmonary hypertension or significant right-sided valvular abnormalities, congestive heart failure, cardiomyopathies, congenital heart disease, pericardial disease, and cardiac transplantation. The validity of this measure rests on the assumption that the prevalence of these clinical indications is low and relatively uniform across the country. However, Malone et al. found that substantial variation in the use of bilateral catheterization persisted among 37 cardiologists at two large community hospitals, even after adjusting for clinical indications.[201]

Another source of potential bias is the large number of catheterizations performed on an outpatient basis.

Construct validity: Does the indicator perform well in identifying true (or actual) quality of care problems?

No studies were found that explicitly address the construct validity of this indicator. Empirical testings show that bilateral catheterization is positively related to coronary artery bypass graft (CABG) and negatively related to laparoscopic cholecystectomy.[202]

Fosters true quality improvement: Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?

Bilateral cardiac catheterization does not generally affect hospital payment; therefore, widespread use of this indicator may lead to less frequent coding when the procedure is performed. A reduction in the rate of bilateral cardiac catheterization may lead to rare, but potentially serious, missed diagnoses (e.g., pulmonary hypertension).

Prior use: Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

Bilateral cardiac catheterization has been widely used as an indicator of quality in the Medicare program and is one of five quality indicators included in the Medicare Quality of Care Report of Surveillance Measures.[203] The success of education and outreach projects suggests that right heart catheterization rates represent an actionable opportunity for quality improvement.

Cardiology/American Heart Association Ad Hoc Task Force on Cardiac Catheterization. Circulation 1991;84(5):2213-47.

[199] Hall M, Popovic J. 1998 summary: National Hospital Discharge Survey. Advance Data from Vital and Health Statistics 2000;316.

[200] Nationwide Inpatient Sample and State Databases. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/data/hcup

[201] Malone ML, Bajwa TK, Battiola RJ, et al. Variation among cardiologists in the utilization of right heart catheterization at time of coronary angiography [see comments]. Cathet Cardiovasc Diagn 1996;37(2):125-30.

[202] Nationwide Inpatient Sample.

[203] Medicare Quality of Care Report of Surveillance Measures. Centers for Medicare and Medicaid Services (formerly Health Care Financing Administration), U.S. Department of Health and Human Services.

5.29 Coronary Artery Bypass Graft Area Rate (IQI 26)

Coronary artery bypass graft (CABG) is performed on patients with coronary artery disease. No ideal rate for CABG has been established.

Relationship to Quality	CABG is an elective procedure that may be overused; therefore, more average rates would represent better quality.
Benchmark	State, regional, or peer group average.
Definition	Number of CABGs per 100,000 population.
Numerator	Number of CABGs in any procedure field. All discharges age 40 years and older. Exclude cases: • MDC 14 (pregnancy, childbirth, and puerperium) • MDC 15 (newborns and other neonates).
Denominator	Population in Metro Area or county, age 40 years or older.
Type of Indicator	Area Level, Utilization Indicator
Empirical Performance	Population Rate (2003): 241.41 per 100,000 population at risk
Empirical Rating	19

Summary of Evidence

CABG is a potentially overused procedure, although several studies have noted that CABG is not often performed for inappropriate indications (under 15%). The risk factors associated with CABG include smoking, hyperlipidemia, and older age, and risk adjustment with demographic data—at a minimum—is recommended. This indicator was designed for use with CABG volume and mortality indicators.

This indicator is measured with very high precision. Substantial and systematic small area variation that is not explained by socio-demographic characteristics has been noted in the literature. Examination of data containing patient residence may aid in identifying the extent to which patients are referred into an area.

Limitations on Use

As an area utilization indicator, CABG is a proxy for actual quality problems. This indicator in particular has unclear construct validity, because CABG does not appear to be performed inappropriately often. Caution should be maintained for CABG rates that are drastically below or above the average or recommended rates.

Details

Face validity: Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?

Most previous studies of small area variation have found relatively high variation in CABG rates, as noted by the systematic component of variation (.758), which compares geographic variability between DRGs after removing random effects.[204] This variation is not explained by population characteristics such as age and sex. No randomized controlled trials have demonstrated that CABG improves clinical outcomes in patients with symptoms less major than three-vessel disease, previous myocardial infarction, or less than strongly positive exercise ECG tests.

Precision: Is there a substantial amount of provider or community level variation that is not attributable to random variation?

Precise estimates of utilization can be generated at the area level; however, random variation may become more problematic for relatively small areas (e.g., ZIP codes) or underpopulated areas (e.g., rural counties). Based on empirical

[204] Gittelsohn A, Powe NR, Small area variations in health care delivery in Maryland. Health Serv Res 1995;30(2):295-317.

evidence, the indicator is moderately precise, with a raw area level mean of 180.4 per 100,000 population and a standard deviation of 571.6.[205] Relative to other indicators, a larger percentage of the variation occurs at the area level, rather than the discharge level. The signal ratio (i.e., the proportion of the total variation that is truly related to systematic differences in area performance rather than random variation) is very high, at 97.3%, indicating that observed differences in area performance very likely represent true differences.

Minimal bias: Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?

The prevalence of coronary artery disease may be related to the age structure of the population and the prevalence of behavioral or physiologic risk factors such as smoking and hyperlipidemia. Although race and demographic factors have significant effects on the likelihood of CABG, previous studies have shown that sociodemographic differences account for very little of the observed variation in CABG rates.[206]

Some differences in CABG rates across areas may be attributable to the referral of rural and other patients from outside the area for surgery; however, such referrals are unlikely to explain a large part of the substantial differences in rates across small geographic areas.

Construct validity: Does the indicator perform well in identifying true (or actual) quality of care problems?

Although most studies have found relatively low rates of inappropriate CABG use, there is some evidence of variation in inappropriate rates across geographic areas. In addition, a larger proportion of bypass surgery procedures is performed for indications in which benefits are uncertain; procedure rates for uncertain indications may also vary substantially across hospitals and areas.

In a follow-up to a New York appropriateness study, a panel of cardiologists found a rate of inappropriate procedure of 6% and a rate of uncertain procedures of 12%.[207] In another study of 12 hospitals, the rate of CABG for inappropriate indications ranged from 0% to 5% across hospitals, and the rate of CABG for uncertain indications ranged from 5% to 8%.[208]

Fosters true quality improvement: Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?

Little evidence exists on whether the use of CABG as a quality indicator might differentially reduce procedures that are inappropriate or of unclear benefit, rather than appropriate procedures.

Prior use: Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

The hospital-based rate of CABG was included in the original HCUP QI indicator set. The area-based rate of CABG is a current indicator in the Dartmouth Atlas.[209]

---

[205] Nationwide Inpatient Sample and State Inpatient Databases. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/data/hcup/

[206] Leape LL, Hilborne LH, Park RE, et al. The appropriateness of use of coronary artery bypass graft surgery in New York state. JAMA 1993;269(6):753-60.

[207] Leape LL, Park RE, Bashore TM, et al. Effect of variability in the interpretation of coronary angiograms on the appropriateness of use of coronary revascularization procedures. American Heart Journal 2000;139(1 Pt 1):106-13.

[208] Leape LL, Hilborne LH, Schwartz JS, et al. The appropriateness of coronary artery bypass graft surgery in academic medical centers. Working Group of the Appropriateness Project of the Academic Medical Center Consortium. Ann Intern Med 1996;125(1):8-18.

[209] Dartmouth Atlas of Health Care, Center for the Evaluative Clinical Sciences at Dartmouth Medical School.

5.30 Percutaneous Transluminal Coronary Angioplasty Area Rate (IQI 27)

Percutaneous transluminal coronary angioplasty (PTCA) is performed on patients with coronary artery disease. No ideal rate for PTCA has been established.

Relationship to Quality	PTCA has been identified as a potentially overused procedure; therefore, more average rates represent better quality care.
Benchmark	State, regional, or peer group average.
Definition	Number of PTCA procedures per 100,000 population.
Numerator	Number of PTCA procedures in any procedure field. All discharges age 40 years and older. Exclude cases: • MDC 14 (pregnancy, childbirth, and puerperium) • MDC 15 (newborns and other neonates)
Denominator	Population in Metro Area or county, age 40 years and older.
Type of Indicator	Area Level, Utilization Indicator
Empirical Performance	Population Rate (2003): 568.29 per 100,000 population at risk
Empirical Rating	19

Summary of Evidence

PTCA is a potentially overused procedure, and rates vary widely and systematically between areas. Patient and physician preferences may play a role in this variation. Clinical factors that are appropriate indications for PTCA may be more prevalent in areas with an older age structure or higher rates of smoking or hyperlipidemia. It is unlikely that these factors would account for all the observed variance.

Empirical evidence shows that risk adjustment by age and sex affects the performance of this indicator; without adequate risk adjustment, areas may be mislabeled as outliers. In addition, examination of data containing patient residence may aid in identifying the extent to which patients are referred into an area.

Limitations on Use

As an area utilization indicator, PTCA is a proxy for actual quality problems. The indicator has unclear construct validity, as high utilization of PTCA has not been shown to necessarily be associated with higher rates of inappropriate utilization. A minor source of bias may be the small number of procedures performed on an outpatient basis. Caution should be maintained for PTCA rates that are drastically below or above the average or recommended rates.

Details

Face validity: Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?

No randomized controlled trials have demonstrated that PTCA improves clinical outcomes in many patients who commonly receive the procedure, and previous studies have documented large differences across hospitals in the likelihood of treatment with PTCA after myocardial infarction and in other clinical settings. Studies on small area variation also found substantial variation in PTCA rates.

Precision: Is there a substantial amount of provider or community level variation that is not attributable to random variation?

Precise estimates of utilization can be generated at the area level; however, random variation may become more problematic for relatively small areas (e.g., ZIP codes) or underpopulated areas (e.g., rural counties). Based on empirical evidence, this indicator is precise, with a raw area level mean of 190.8 per 100,000 population and a standard deviation of 455.6.[210]

[210] Nationwide Inpatient Sample and State Inpatient Databases. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/data/hcup/

Relative to other indicators, a higher percentage of the variation occurs at the area level, rather than the discharge level. The signal ratio (i.e., the proportion of the total variation that is truly related to systematic differences in area performance rather than random variation) is very high, at 97.3%, indicating that observed differences in area performance very likely represent true differences.

Minimal bias: Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?

Little evidence exists on the extent to which area differences in socioeconomic and clinical characteristics may explain area differences in PTCA rates, although large variations in rates across small geographic areas suggest that population characteristics are unlikely to explain most of the differences.[211]

Construct validity: Does the indicator perform well in identifying true (or actual) quality of care problems?

For this indicator to perform well in identifying true quality of care problems, there must be evidence of significant inappropriate use in population-based studies, as well as substantial variation in the rate of inappropriate use across providers or small areas. In a study of seven Swedish heart centers, 38.3% of all PTCA procedures were performed for inappropriate indications and 30% for uncertain indications.[212] In a follow-up study of a coronary angiography study conducted in New York, a panel of cardiologists found the rate for inappropriate indications was 12% and the rate of procedures performed for uncertain indications was 27%.[213]

Fosters true quality improvement: Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?

Providers might engage in practices such as miscoding cases or recruiting patient groups that are known to have increased risk of coronary artery disease to achieve more favorable quality assessment results. Instead of serving as quality assessments, patients and their providers might use the results of appropriateness studies to spark questions and discussion about coronary artery disease, the patient's specific indications, and the treatment that poses the least risk to the patient.[214]

Prior use: Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

The area-based rate of PTCA is a current indicator in the Dartmouth Atlas.[215]

[211] Ziskind AA, Lauer MA, Bishop G, et al. Assessing the appropriateness of coronary revascularization: the University of Maryland Revascularization Appropriateness Score (RAS) and its comparison to RAND expert panel ratings and American College of Cardiology/American Heart Association guidelines with regard to assigned appropriateness rating and ability to predict outcome. Clin Cardiol 1999;22(2):67-76.
[212] Bernstein SJ, Brorsson B, Aberg T, et al. Appropriateness of referral of coronary angiography patients in Sweden. SECOR/SBU Project Group. Heart 1999;81(5):470-7.
[213] Leape LL, Park RE, Bashore TM, et al. Effect of variability in the interpretation of coronary angiograms on the appropriateness of use of coronary revascularization procedures. American Heart Journal 2000;139(1 Pt 1):106-13.
[214] Hilborne LH, Leape LL, Bernstein SJ, et al. The appropriateness of use of percutaneous transluminal coronary angioplasty in New York state. JAMA 1993;269(6):761-5.
[215] Dartmouth Atlas of Health Care, Center for the Evaluative Clinical Sciences at Dartmouth Medical School.

5.31 Hysterectomy Area Rate (IQI 28)

Hysterectomy is performed on patients with a number of indications, such as recurrent uterine bleeding, chronic pelvic pain, or menopause, usually in some combination. No ideal rate for hysterectomy has been established.

Relationship to Quality	Hysterectomy has been identified as a potentially overused procedure; therefore, more average rates represent better quality care.
Benchmark	State, regional, or peer group average.
Definition	Number of hysterectomies per 100,000 population.
Numerator	Number of hysterectomies in any procedure field. All discharges of females age 18 years and older. Exclude cases: • with genital cancer or pelvic or lower abdominal trauma in any diagnosis field • MDC 14 (pregnancy, childbirth, and puerperium) • MDC 15 (newborns and other neonates).
Denominator	Female population in Metro Area or county age 18 years and older.
Type of Indicator	Area Level, Utilization Indicator
Empirical Performance	Population Rate (2003): 464.34 per 100,000 population at risk
Empirical Rating	22

Summary of Evidence

Hysterectomy is a potentially overused procedure. Population rates have been shown to vary systematically by small geographic area; however, patient and physician preference may play a role in the choice to have a hysterectomy, which in turn may affect area rates. Examination of data containing patient residence may aid in identifying the extent to which patients are referred into an area.

This indicator is not expected to be substantially biased, because it is unlikely that appropriate indications for hysterectomy would vary systematically by area. However, risk adjustment with age is recommended. Although the ideal rate for hysterectomy has not been established, several studies have noted relatively high rates of inappropriate indicators for surgery (16-70%).

Limitations on Use

As an area utilization indicator, hysterectomy is a proxy for actual quality problems. The indicator has unclear construct validity, as high utilization of hysterectomy has not been shown to necessarily be associated with higher rates of inappropriate utilization. Additional clinical risk adjustment, such as for parity, may be desirable. Caution should be maintained for hysterectomy rates that are drastically below or above the average or recommended rates.

Details

Face validity: Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?

No randomized controlled trials have demonstrated that hysterectomy improves outcomes in patients with uncertain clinical indications, including persistent or recurrent abnormal bleeding, pain, adnexal mass, limited hormonal therapy, and premenopausal age.

Small area variation has been noted in the literature on hysterectomy rates.[216]

Precision: Is there a substantial amount of provider or community level variation that is not attributable to random variation?

Precise estimates of utilization can be generated at the area level; however, random variation may become more problematic for relatively

[216]Gittlesohn A, Powe NR. Small area variations in health care delivery in Maryland. Health Serv Res 1995;30(2):295-317.

small areas (e.g., ZIP codes) or underpopulated areas (e.g., rural counties). Based on empirical evidence, this indicator is precise, with a raw area level rate of 419.4 per 100,000 population and a substantial standard deviation of 323.3.[217]

Relative to other indicators, a higher percentage of the variation occurs at the area level, rather than the discharge level. The signal ratio (i.e., the proportion of the total variation that is truly related to systematic differences in area performance rather than random variation) is very high, at 93.6%, indicating that observed differences in area performance likely represent true differences.

Minimal bias: Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?

Utilization rates standardized at the area level (e.g., adult population of the county or standard metro area) may be biased by differences in the prevalence of those indications that warrant the procedure. The prevalence of these indications may, in turn, be related to the age structure of the population and the prevalence of behavioral or physiologic risk factors. In a study of seven managed care organizations, older women were more likely than younger women to have received a hysterectomy for appropriate reasons.[218]

Construct validity: Does the indicator perform well in identifying true (or actual) quality of care problems?

For this indicator to perform well in identifying true quality of care problems, there must be evidence of significant inappropriate use in population-based studies, as well as substantial variation in the rate of inappropriate use across providers or small areas. In a random sample of 642 hysterectomies, 16% of procedures were inappropriate based on patient indications, and 25% were uncertain.[219] Another study found a 70% rate of overall inappropriate indications, varying from 45% to 100% across diagnoses indicative of hysterectomy.[220]

Fosters true quality improvement: Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?

Little evidence exists on whether hysterectomy as a quality indicator might reduce appropriate as well as inappropriate hysterectomies, or the extent to which overall hysterectomy rates are correlated with inappropriate hysterectomy rates.

Prior use: Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

The hospital-based rate of hysterectomy was included in the original HCUP QI indicator set. The area-based rate of hysterectomy is a current indicator in the Dartmouth Atlas.[221]

[217] Nationwide Inpatient Sample and State Inpatient Databases. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/data/hcup/

[218] Bernstein SJ, McGlynn EA, Siu AL, et al. The appropriateness of hysterectomy. A comparison of care in seven health plans. Health Maintenance Organization Quality of Care Consortium [see comments]. JAMA 1993;269(18):2398-402.

[219] Bernstein et al., 1993.

[220] Broder MS, Kanouse DE, Mittman BS, et al. The appropriateness of recommendations for hysterectomy. Obstet Gynecol 2000;95(2):199-205.

[221] Dartmouth Atlas of Health Care, Center for the Evaluative Clinical Sciences at Dartmouth Medical School.

5.32 Laminectomy or Spinal Fusion Area Rate (IQI 29)

Laminectomy is performed on patients with a herniated disc or spinal stenosis. No ideal rate for laminectomy has been established.

Relationship to Quality	Laminectomy has been identified as a potentially overused procedure; therefore, more average rates represent better quality care.
Benchmark	State, regional, or peer group average.
Definition	Number of laminectomies or spinal fusions per 100,000 population.
Numerator	Number of laminectomies or spinal fusions in any procedure field. All discharges age 18 years and older. Exclude cases: • MDC 14 (pregnancy, childbirth, and puerperium) • MDC 15 (newborns and other neonates).
Denominator	Population in Metro Area or county, age 18 years and older.
Type of Indicator	Area Level, Utilization Indicator
Empirical Performance	Population Rate (2003): 252.77 per 100,000 population at risk
Empirical Rating	20

Summary of Evidence

Laminectomy, which is a potentially overused procedure, has been shown to vary widely and systematically between areas. Patient and physician preference may play a role in the decision to have a laminectomy, which may in turn affect area rates.

Empirical analysis suggests that performance is not highly influenced by the demographic breakdown of the population. Without adequate risk adjustment for age and sex, areas may be mislabeled as outliers. Although the ideal rate for laminectomy has not been established, several studies have noted relatively high rates of inappropriate procedures (23-38%).

High area rates may not take into account that some patients are referred into an area hospital from a different area. Examination of data with patient residence can help in determining the extent to which patients are referred into the area.

Limitations on Use

As an area utilization indicator, laminectomy is a proxy for actual quality problems. The indicator has unclear construct validity, as high utilization of laminectomy has not been shown to necessarily be associated with higher rates of inappropriate utilization. Caution should be maintained for laminectomy rates that are drastically below or above the average or recommended rates.

Details

Face validity: Does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?

No randomized controlled trials have demonstrated that laminectomy improves outcomes in patients with uncertain clinical indications, including minor neurological findings, lengthy restricted activity, and equivocal imaging for discal hernia or spinal stenosis.

Prior research on small area variation has found relatively high variation in laminectomy rates.[222] Larequi-Lauber et al. report that the use of back surgery in the United States varies from one area to another by as much as 15-fold.[223] This high variation was not explained by population characteristics such as age and sex.

[222] Gittlesohn A, Powe NR. Small area variations in health care delivery in Maryland. Health Serv Res 1995;30(2):295-317.
[223] Larequi-Lauber T, Vader JP, Burnand B, et al. Appropriateness of indications for surgery of lumbar disc hernia and spinal stenosis. Spine 1997;22(2):203-9.

Precision: Is there a substantial amount of provider or community level variation that is not attributable to random variation?

Precise estimates of utilization can be generated at the area level; however, random variation may become more problematic for relatively small areas (e.g., ZIP codes) or underpopulated areas (e.g., rural counties). Based on empirical evidence, this indicator is moderately precise, with a raw area level mean of 139.0 per 100,000 population and a standard deviation of 347.5.[224]

Relative to other indicators, a higher percentage of the variation occurs at the area level, rather than the discharge level. The signal ratio (i.e., the proportion of the total variation that is truly related to systematic differences in area performance rather than random variation) is very high, at 96.7%, indicating that observed differences in area performance very likely represent true differences.

Minimal bias: Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?

Utilization rates standardized at the area level (e.g., county or metro area) may be biased by differences in the prevalence of herniated disc or spinal stenosis, which may in turn be related to the age structure of the population and the prevalence of behavioral or physiologic risk factors. However, studies have shown that sociodemographic differences and other measurable population characteristics account for very little or none of the observed variation in laminectomy rates.[225]

Construct validity: Does the indicator perform well in identifying true (or actual) quality of care problems?

For this indicator to perform well in identifying true quality of care problems, there must be evidence of significant inappropriate use in population-based studies, as well as substantial variation in the rate of inappropriate use across providers or small areas. In an assessment of cases at one Swiss hospital, 23% of patients received surgical treatment for herniated discs for inappropriate reasons and 29% received surgical treatment for uncertain indications.[226] In another study of teaching hospital patients undergoing surgery for herniated disc or spinal stenosis, 38% of surgeries were performed for inappropriate indications.

Fosters true quality improvement: Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?

Little evidence exists on whether use of laminectomy as a quality indicator would lead to less performance of laminectomies for inappropriate or uncertain indications without reducing the use of laminectomy for appropriate indications.

Prior use: Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

The hospital-based rate of laminectomy was included in the original HCUP QI indicator set. The area-based rate of laminectomy is a current indicator in the Dartmouth Atlas.[227]

[224] Nationwide Inpatient Sample and State Inpatient Databases. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/data/hcup

[225] Barron M, Kazandjian VA. Geographic variation in lumbar diskectomy: a protocol for evaluation. QRB Qual Rev Bull 1992;18(3):98-107.

[226] Porchet F, Vader JP, Larequi-Lauber T, et al. The assessment of appropriate indications for laminectomy. J Bone Joint Surg Br 1999;81(2):234-9.

[227] Dartmouth Atlas of Health Care, Center for the Evaluative Clinical Sciences at Dartmouth Medical School.

6.0 Using Different Types of QI Rates

When should you use the observed, expected, risk adjusted, and/or smoothed rates generated by the AHRQ QI software? Here are some guidelines.

If the user's primary interest is to identify cases for further follow-up and quality improvement, then the observed rate would help to identify them. The observed rate is the raw rate generated by the QI software from the data the user provided. Areas for improvement can be identified by the magnitude of the observed rate compared to available benchmarks and/or by the number of patients impacted.

Additional breakdowns by the default patient characteristics used in stratified rates (e.g., age, gender, or payer) can further identify the target population. Target populations can also be identified by user-defined patient characteristics supplemented to the case/discharge level flags. Trend data can be used to measure change in the rate over time.

Another approach to identify areas to focus on is to compare the observed and expected rates. The expected rate is the rate the provider would have if it performed the same as the reference population given the provider's actual case-mix (e.g., age, gender, DRG, and comorbidity categories).

If the observed rate is higher than the expected rate (i.e., the ratio of observed/expected is greater than 1.0, or observed minus expected is positive), then the implication is that the provider performed worse than the reference population for that particular indicator. Users may want to focus on these indicators for quality improvement.

If the observed rate is lower than the expected rate (i.e., the ratio of observed/expected is less than 1.0, or observed minus expected is negative), then the implication is that the provider performed better than the reference population. Users may want to focus on these indicators for identifying best practices.

Users can also compare the expected rate to the population rate reported in the detailed evidence section of the IQI, PQI, or PSI Guide to determine how their case-mix compares to the reference population. If the population rate is higher than the expected rate, then the provider's case-mix is less severe than the reference population. If the population rate is lower than the expected rate, then the provider's case-mix is more severe than the reference population.

We use this difference between the population rate and the expected rate to "adjust" the observed rate to account for the difference between the case-mix of the reference population and the provider's case-mix. This is the provider's risk-adjusted rate.

If the provider has a less severe case-mix, then the adjustment is positive (population rate > expected rate) and the risk-adjusted rate is higher than the observed rate. If the provider has a more severe case-mix, then the adjustment is negative (population rate < expected rate) and the risk-adjusted rate is lower than the observed rate. The risk-adjusted rate is the rate the provider would have if it had the same case-mix as the reference population given the provider's actual performance.

Finally, users can compare the risk-adjusted rate to the smoothed or "reliability-adjusted" rate to determine whether this difference between the risk-adjusted rate and reference population rate is likely to remain in the next measurement period. Smoothed rates are weighted averages of the population rate and the risk-adjusted rate, where the weight reflects the reliability of the provider's risk-adjusted rate.

A ratio of (smoothed rate - population rate) / (risk-adjusted rate - population rate) greater than 0.80 suggests that the difference is likely to persist (whether the difference is positive or negative). A ratio less than 0.80 suggests that the difference may be due in part to random differences in patient characteristics (patient characteristics that are not observed and controlled for in the risk-adjustment model). In general, users may want to focus on areas where the differences are more likely to persist.

7.0 References

Al-Mufti R, McCarthy A, Fisk NM. Obstetricians' personal choice and mode of delivery [letter] [see comments]. Lancet 1996;347(9000):544.

American Heart Association. Heart Attack and Stroke Facts: 1996 Statistical Supplement. Dallas, TX: American Heart Association; 1996.

Amundsen S, Skjaerven R, Trippestad A, et al. 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 Chir Scand 1990;156(4):323-7; discussion 327-8.

Andrew MH, Roty AR, Jr. Incidental appendectomy with cholecystectomy: is the increased risk justified? Am Surg 1987;53(10):553-7.

Aron DC, Harper DL, Shepardson LB, et al. Impact of risk-adjusting cesarean delivery rates when reporting hospital performance. JAMA 1998;279(24):1968-72.

Balit JL, Dooley SL, Peaceman AN. Risk adjustment for interhospital comparison of primary cesarean rates. Obstet Gynecol 1999;93(6):1025-30.

Ball JK, Elixhauser A, Johantgen M, et al. *HCUP Quality Indicators, Methods, Version 1.1: Outcome, Utilization, and Access Measures for Quality Improvement*. (AHCPR Publication No. 98-0035). Healthcare Cost and Utilization project (HCUP-3) Research notes: Rockville, MD: Agency for Health Care Policy and Research, 1998.

Barron M, Kazandjian VA. Geographic variation in lumbar diskectomy: a protocol for evaluation. QRB Qual Rev Bull 1992;18(3):98-107.

Begg CB, Cramer LD, Hoskins WJ, et al. Impact of hospital volume on operative mortality for major cancer surgery. JAMA 1998;280(20):1747-51.

Berkowitz GS, Fiarman GS, Mojica MA, et al. Effect of physician characteristics on the cesarean birth rate [see comments]. Am J Obstet Gynecol 1989;161(1):146-9.

Bernstein SJ, Brorsson B, Aberg T, et al. Appropriateness of referral of coronary angiography patients in Sweden. SECOR/SBU Project Group. Heart 1999;81(5):470-7.

Bernstein SJ, McGlynn EA, Siu AL, et al. The appropriateness of hysterectomy. A comparison of care in seven health plans. Health Maintenance Organization Quality of Care Consortium [see comments]. JAMA 1993;269(18):2398-402.

Bickell NA, Zdeb MS, Applegate MS, et al. Effect of external peer review on cesarean delivery rates: a statewide program. Obstet Gynecol 1996;87(5 Pt 1):664-7.

Biller J, Feinberg WM, Castaldo JE, et al. Guidelines for carotid endarterectomy: a statement of healthcare professionals from a Special Writing Group of the Stroke Council, American Heart Association. Circulation 1998;97(5):501-9.

Birkmeyer JD, Finlayson SR, Tosteson AN, et al. Effect of hospital volume on in-hospital mortality with pancreaticoduodenectomy. Surgery 1999;125(3):250-6.

Broder MS, Kanouse DE, Mittman BS, et al. The appropriateness of recommendations for hysterectomy. Obstet Gynecol 2000;95(2):199-205.

Brown RD, Whisnant JP, Sicks JD, et al. Stroke incidence, prevalence, and survival: secular trends in Rochester, Minnesota, through 1989. Stroke 1996;27(3):373-80.

Cebul RD, Snow RJ, Pine R, et al. Indications, outcomes, and provider volumes for carotid endarterectomy. JAMA 1998;279(16):1282-7.

Centers for Disease Control and Prevention. Report of Final Mortality Statistics, 1996. Volume 47, Number 9. Available at http://www.cdc.gov/nchs/.

Cook DJ, Guyatt GH, Salena BJ, et al. Endoscopic therapy for acute nonvariceal upper gastrointestinal hemorrhage: a meta-analysis. Gastroenterology 1992;102(1):139-48.

Cooper GS, Chak A, Way LE, et al. Early endoscopy in upper gastrointestinal hemorrhage: associations with recurrent bleeding, surgery, and length of hospital stay. Gastrointest Endosc 1999;49(2):145-52.

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. J Vasc Surg 1998;28(3):413-20.

Dartmouth Atlas of Health Care, Center for the Evaluative Clinical Sciences at Dartmouth Medical School, 1999.

Davis RB, Iezzoni LI, Phillips RS, et al. Predicting in-hospital mortality. The importance of functional status information. Med Care 1995;33(9):906-21.

Dudley RA, Johansen KL, Brand R, et al. Selective referral to high-volume hospitals: estimating potentially avoidable deaths. JAMA 2000;283(9):1159-66.

Eagle KA, Guyton RA, Davidoff R, et al. ACC/AHA Guidelines for Coronary Artery Bypass Graft Surgery: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Revise the 1991 Guidelines for Coronary Artery Bypass Graft Surgery). American College of Cardiology/American Heart Association. J Am Coll Cardiol 1999;34(4):1262-347.

Escarce JJ, Chen W, Schwartz JS. Falling cholecystectomy thresholds since the introduction of laparoscopic cholecystectomy. JAMA 1995;273(20):1581-5.

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.

Feinberg WM. Guidelines for the management of transient ischemic attacks. Ad Hoc Committee on Guidelines for the Management of Transient Ischemic Attacks of the Stroke Council, American Heart Association, Heart Dis Stroke 1994;3(5):275-83.

Fine MJ, Smith MA, Carson CA, et al. Prognosis and outcomes of patients with community-acquired pneumonia. A meta-analysis. JAMA 1996;275(2):134-41.

Fisher KS, Ross DS. Guidelines for therapeutic decision in incidental appendectomy. Surg Gynecol Obstet 1990;171(1):95-8.

Forsen L, Sogaard AJ, Meyer HE, et al. Survival after hip fracture: short- and long-term excess mortality according to age and gender. Osteoporos Int 1999;10(1):73-8.

Gentles TL, Mayer JE, Jr., Gauvreau K, et al. Fontan operation in 500 consecutive patients: factors influencing early and late outcome. J Thorac Cardiovasc Surg 1997;114(3):376-91.

Ghali WA, Ash AS, Hall RE, et al. Statewide quality improvement initiatives and mortality after cardiac surgery. JAMA 1997;277(5):379-82.

Gittelsohn A, Powe NR, Small area variations in health care delivery in Maryland. Health Serv Res 1995;30(2):295-317.

Glasgow RE, Mulvihill SJ. Hospital volume influences outcome in patients undergoing pancreatic resection for cancer. West J Med 1996;165(5):294-300.

Gleason PP, Heehan TP, Fine JM, et al. Associations between initial antimicrobial therapy and medical outcomes for hospitalized elderly patients with pneumonia. Arch Intern Med 1999;159(21):2562-72.

Goff, DC, Jr., Pandey DK, Chan FA, et al. Congestive heart failure in the United States: is there more than meets the I(CD code)? The Corpus Christi Heart Project. Arch Intern Med 2000;160(2):197-202.

Gordan 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.

Goyert GL, Bottoms FS, Treadwell MC, et al. The physician factor in cesarean birth rates [see comments]. N Engl J Med 1989;320(11):706-9.

Green DC, Moore JM, Adams MM, et al. Are we underestimating rates of vaginal birth after previous cesarean birth? The validity of delivery methods from birth certificates. Am J Epidemiol 1998;147(6):581-6.

Haas JS, Udvarhelyi S, Epstein AM. The effect of health coverage for uninsured pregnant women on maternal health and the use of cesarean section [see comments]. JAMA 1993;270(1):61-4.

Hall M, Popovic J. 1998 summary: National Hospital Discharge Survey. Advance Data from Vital and Health Statistics 2000;316.

Hannan EL, Racz M, Kavey RE, et al. Pediatric cardiac surgery: the effect of hospital and surgeon volume on in-hospital mortality. Pediatrics 1998;101(6):963-9.

Hannan EL, Popp AJ, Tranmer B, et al. Relationship between provider volume and mortality for carotid endarterectomies in New York state. Stroke 1998;29(11):2292-7.

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 1992;27(4):517-42.

Hannan EL, Racz M, Ryan TJ, et al. Coronary angioplasty volume-outcome relationships for hospitals and cardiologists. JAMA 1997;277(11):892-8.

Hannan EL, Kilburn H, Jr., Bernard H, et al. Coronary artery bypass surgery: the relationship between inhospital mortality rate and surgical volume after controlling for clinical risk factors. Med Care 1991;29(11):1094-107.

Hannan EL, Racz M, Kavey RE, et al. Pediatric cardiac surgery: the effect of hospital and surgeon volume on in-hospital mortality. Pediatrics 1998;101(6):963-9.

Hannan EL, Kilburn H Jr., Racz M, et al. Improving the outcomes of coronary artery bypass surgery in New York state. JAMA 1994;271(10):761-6.

Hannan EL, Siu AL, Kumar D, et al. Assessment of coronary artery bypass graft surgery performance in New York. Is there a bias against taking high-risk patients? Med Care 1997;35(1):49-56.

Hay JA, Lyubashevsky E, Elashoff J, et al. Upper gastrointestinal hemorrhage clinical guideline determining the optimal hospital length of stay. Am J Med 1996;100(3):313-22.

HCUPnet. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://hcup.ahrq.gov/HCUPnet.asp.

Healthy People 2010. Office of Disease Prevention and Health Promotion, U.S. Department of Health and Human Services, Washington, DC: U.S. Government Printing Office.

Higgins TL, Estafanous FG, Loop FD, et al. Stratification of morbidity and mortality outcome by preoperative risk factors in coronary artery bypass patients. A clinical severity score. JAMA 1992;267(17):2344-8.

Hilborne LH, Leape LL, Bernstein SJ, et al. The appropriateness of use of percutaneous transluminal coronary angioplasty in New York state. JAMA 1993;269(6):761-5.

Hirshfeld JW, Jr., Ellis SG, Faxon DP. Recommendations for the assessment and maintenance of proficiency in coronary interventional procedures: Statement of the American College of Cardiology. J Am Coll Cardiol 1998;31(3):722-43.

Hoyert DL, Kochanek KD, Murphy SL. Deaths: final data for 1997. Natl Vital Stat Rep 1999;47(19):1-104.

http://www.cms.hhs.gov/statistics/nhe/projections-2002/t2.asp: Table 2: National Health Expenditure Amounts, and Average Annual Percent Change by Type of Expenditure: Selected Calendar Years 1980-2012.

Impact: Case Studies Notebook – Documented Impact and Use of AHRQ's Research. Compiled by Division of Public Affairs, Office of Health Care Information, Agency for Healthcare Research and Quality.

Institute of Medicine Division of Health Care Services. Medicare: a strategy for quality assurance. Washington, DC: National Academy Press; 1990.

Institute of Medicine. To Err is Human: Building a Safer Health System. Kohn LT, Corrigan JM, Donaldson MS (eds.) Washington DC: National Academy Press, 2000.

Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Committee of Quality of Care in America. Washington DC: National Academy Press, 2001.

Jatzko GR, Lisborg PH, Pertl AM, et al. Multivariate comparison of complications after laparoscopic cholecystectomy and open cholecystectomy. Ann Surg 1995;221(4):381-6.

Jenkins KJ, Newburger JW, Lock JE, et al. In-hospital mortality for surgical repair of congenital heart defects: preliminary observations of variation by hospital caseload. Pediatrics 1995;95(3):323-30.

Kahn KL, Rogers WH, Rubenstein LV, et al. Measuring quality of care with explicit process criteria before and after implementation of the DRG-based prospective payment system. JAMA 1990;264(15):1969-73.

Kalra L, Perez I, Smithard DG, et al. Does prior use of aspirin affect outcome in ischemic stroke? Am J Med 2000;108(3):205-9.

Kantonen I, Lepantalo M, Brommels M, et al. Mortality in ruptured abdominal aortic aneurysms. The Finnvasc Study Group. Eur J Vasc Endovasc Surg 1999;17(3):208-12.

Karp, HR, Flanders WD, Shipp CC, et al. Carotid endarterectomy among Medicare beneficiaries: a statewide evaluation of appropriateness and outcome. Stroke 1998;29(1):46-52.

Kazandjian VA, Lied TR. Cesarean section rates: effects of participation in a performance measurement project. Jt Comm J Qual Improv 1998;24(4):187-96.

Kazmers A, Jacobs L, Perkins A, et al. Abdominal aortic aneurysm repair in Veterans Affairs medical centers. J Vasc Surg 1996;23(2):191-200.

Kern JH, Hayes CJ, Michler RE, et al. Survival and risk factor analysis for the Norwood procedure for hypoplastic left heart syndrome. Am J Cardiol 1997;80(2):170-4.

Knott-Craig CJ, Danielson GK, Schaff HV, et al. The modified Fontan operation. An analysis of risk factors for early postoperative death or takedown in 702 consecutive patients from one institution. J Thorac Cardiovasc Surg 1995;109(6):1237-43.

Kozak LJ, Lawrence L. National Hospital Discharge Survey: annual summary, 1997. Vital Health Stat 13 1999(144):i-iv, 1-46.

Kreder HF, Williams JI, Jaglal S, et al. 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.

Landon B, Iezzoni LI, Ash AS, et al. Judging hospitals by severity-adjusted mortality rates: the case of CABG surgery. Inquiry 1996;33(2):155-66.

Lanzino G, Kassell NF, Germanson TP, et al. Age and outcome after aneurysmal subarachnoid hemorrhage: why do older patients fare worse? J Neurosurg 1996;85(3):410-8.

Larequi-Lauber T, Vader JP, Burnand B, et al. Appropriateness of indications for surgery of lumbar disc hernia and spinal stenosis. Spine 1997;22(2):203-9.

Lavernia CJ, Guzman JF. Relationship of surgical volume to short-term mortality, morbidity, and hospital charges in arthroplasty. J Arthroplasty 1995;10(2):133-40.

Leape LL, Hilborne LH, Schwartz JS, et al. The appropriateness of coronary artery bypass graft surgery in academic medical centers. Working Group of the Appropriateness Project of the Academic Medical Center Consortium. Ann Intern Med 1996;125(1):8-18.

Leape LL, Park RE, Bashore TM, et al. Effect of variability in the interpretation of coronary angiograms on the appropriateness of use of coronary revascularization procedures. American Heart Journal 2000;139(1 Pt 1):106-13.

Leape LL, Hilborne LH, Park RE, et al. The appropriateness of use of coronary artery bypass graft surgery in New York state. JAMA 1993;269(6):753-60.

Lieberman MD, Kilburn H, Lindsey M, et al. Relation of perioperative deaths to hospital volume among patients undergoing pancreatic resection for malignancy. Ann Surg 1995;222(5):638-45.

Localio AR, Hamory BH, Fisher AC, et al. The public release of hospital and physician mortality data in Pennsylvania. A case study. Med Care 1997;35(3):272-286.

Luft HS, Hunt SS, Maerki SC. The volume-outcome relationship: practice-makes-perfect or selective-referral patterns? Health Serv Res 1987;22(2):157-82.

MacIntyre K, Capewell IS, Stewart S, et al. Evidence of improving prognosis in heart failure: trends in case fatality in 66,547 patients hospitalized between 1986 and 1995 [see comments]. Circulation 2000;102(10):1126-31.

Malone ML, Bajwa TK, Battiola RJ, et al. Variation among cardiologists in the utilization of right heart catheterization at time of coronary angiography [see comments]. Cathet Cardiovasc Diagn 1996;37(2):125-30.

Manheim LM, Sohn MW, Feinglass J, et al. Hospital vascular surgery volume and procedure mortality rates in California, 1982-1994. J Vasc Surg 1998;28(1):45-46.

Markowitz JS, Pashko S, Gutterman EM, et al. Death rates among patients hospitalized with community-acquired pneumonia: a reexamination with data from three states. Am J Public Health 1996;86(8 Pt 1):1152-4.

McMahon AJ, Russell IT, Baxter JN, et al. Laparoscopic and minilaparotomy cholecystectomy: a randomised trial [see comment]. Lancet 1994;343(8890):135-8.

McMahon AF, Russell IT, Ramsay G, et al. Laparoscopic and minilaparotomy cholecystectomy: a randomized trial comparing postoperative pain and pulmonary function. Surgery 1994;115(5):533-9.

Medicare Quality of Care Report of Surveillance Measures. Centers for Medicare and Medicaid Services (formerly Health Care Financing Administration), U.S. Department of Health and Human Services.

Meehan TP, Hennen J, Radford MJ, et al. Process and outcome of care for acute myocardial infarction among Medicare beneficiaries in Connecticut: a quality improvement demonstration project. Ann Intern Med 1995;122(12):928-36.

Meehan TP, Fine MJ, Krumholz HM, et al. Quality of care, process, and outcomes in elderly patients with pneumonia. JAMA 1997;278(23):2080-4.

Menard MK. Cesarean delivery rates in the United States. The 1990s. Obstet Gynecol Clin North Am 1999;26(2):275-86.

Myers AH, Robinson EG, Van Natta ML, et al. Hip fractures among the elderly: factors associated with in-hospital mortality. Am J Epidemiol 1991;134(10):1128-37.

Nationwide Inpatient Sample and State Inpatient Databases. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/data/hcup

Neulander EZ, Hawke CK, Soloway MS. Incidental appendectomy during radical cystectomy: an interdepartmental survey and review of the literature. Urology 2000;56(2):241-4.

Ni H, Hershberger FE. Was the decreasing trend in hospital mortality from heart failure attributable to improved hospital care? The Oregon experience, 1991-1995. Am J Manag Care 1999;5(9):1105-15.

Nockerts SR, Detmer DE, Fryback, DG. Incidental appendectomy in the elderly? No. Surgery 1980;88(2):301-6.

Ottino G, Bergerone S, Di Leo M, et al. Aortocoronary bypass results: a discriminant multivariate analysis of risk factors of operative mortality. J Cardiovasc Surg (Torino) 1990;31(1):20-5.

Owings, MF, Lawrence L. Detailed diagnoses and procedures, National Hospital Discharge Survey, 1997. Vital Health Stat 13 199(145):1-157.

Pacific Business Group on Health. (http://www.pbgh.org/)

Patti MG, Corvera CU, Glasgow RE, et al. A hospital's annual rate of esophagectomy influences the operative mortality rate. J Gastrointest Surg 1998;2(2):186-92.

Pearce WH, Parker MA, Feinglass J, et al. The importance of surgeon volume and training in outcomes for vascular surgical procedures. J Vasc Surg 1999;29(5):768-76.

Pepine CJ, Allen HD, Bashore TM, et al. ACC/AHA guidelines for cardiac catheterization and cardiac catheterization laboratories. American College of Cardiology/American Heart Association Ad Hoc Task Force on Cardiac Catheterization. Circulation 1991;84(5):2213-47.

Perez JV, Warwick DJ, Case CP, et al. Death after proximal femoral fracture—an autopsy study. Injury 1995;26(4):237-40.

Peterson ED, DeLong ER, Jollis JG, et al. Public reporting of surgical mortality: a survey of new York State cardiothoracic surgeons. Ann Thorac surg 1999;68(4):1195-200; discussion 12-1-2.

Pilcher DB, Davis JH, Ashikaga T, et al. Treatment of abdominal aortic aneurysm in an entire state over 7½ years. Am J Surg 1980;139(4):487-94.

Popovic JR, Kozak LJ. National hospital discharge survey: annual summary, 1998 [In Process Citation]. Vital Health Stat 13 2000(148):1-194.

Porchet F, Vader JP, Larequi-Lauber T, et al. The assessment of appropriate indications for laminectomy. J Bone Joint Surg Br 1999;81(2):234-9.

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.

Psaty BM, Boineau R, Kuller LH, et al. The potential costs of upcoding for heart failure in the United States. Am J Cardiol 1999;84(1):108-9, A9.

Ritchie JL, Maynard C, Chapko MK, et al. 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.

Roberts RG, Bell HS, Wall EM, et al. Trial of labor or repeated cesarean section. The woman's choice. Arch Fam Med 1997;6(2):120-5.

Rockall TA, Logan RF, Devlin HB, et al. Variation in outcome after acute upper gastrointestinal haemorrhage. The National Audit of Acute Upper Gastrointestinal Haemorrhage. Lancet 1995;346(8971):346-50.

Rosenthal GE, Baker DW, Norris DG, et al. Relationships between in-hospital and 30-day standardized hospital mortality: implications for profiling hospitals. Health Serv Res 2000;34(7):1449-68.

Rutledge R, Oller DW, Meyer AA, et al. A statewide, population-based time-series analysis of the outcome of ruptured abdominal aortic aneurysm. Ann Surg 1996;223(5):492-502.

Ryan TJ, Bauman WB, Kennedy JW, et al. Guidelines for percutaneous transluminal coronary angioplasty. A report of the American Heart Association/American College of Cardiology Task Force on Assessment of Diagnostic and Therapeutic Cardiovascular Procedures (Committee on Percutaneous Transluminal Coronary Angioplasty). Circulation 1993;88(6):2987-3007.

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(3):890-911.

Sachs BP, Kobelin C, Castro MA, et al. The risks of lowering the cesarean-delivery rate. N Engl J Med 1999;340(1):54-7.

Samsa GP, Bian J, Lipscomb J, et al. Epidemiology of recurrent cerebral infarction: a Medicare claims-based comparison of first and recurrent strokes on 2-year survival and cost. Stroke 1999;30(2):338-49.

Second Report of the California Hospitals Outcomes Project, May 1996. Acute Myocardial Infarction. Sacramento, CA: Office of Statewide Health Planning and Development; 1996.

Simunovic M, To T, Theriault M, et al. Relation between hospital surgical volume and outcome for pancreatic resection for neoplasm in a publicly funded health care system [see comments]. Cmaj 1999;160(5):643-8.

Smith, WM. Epidemiology of congestive heart failure. Am J Cardiol 1985;55(2):3A-8A.

Sollano JA, Gelijns AC, Moskowitz AJ et al. Volume-outcome relationships on cardiovascular operations: New York State, 1990-1995. J Thorac Cardiovasc Surg 1999;117(3):419-28.

Soloman 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.

Southern Surgeons Club. A prospective analysis of 1518 laparoscopic cholecystectomies. NEJM 1991;324:1073-1078.

Stachniak JB, Layon AJ, Day AL, et al. Craniotomy for intracranial aneurysm and subarachnoid hemorrhage. Is course, cost, or outcome affected by age? Stroke 1996;27(2):276-81.

Stafford RS. The impact of nonclinical factors on repeat cesarean section [see comments]. JAMA 1991;265(1):59-63.

Stafford RS, Sullivan SD, Gardner LB. Trends in cesarean section use in California, 1983 to 1990. Am J Obstet Gynecol 1993;168(4):1297-302.

Strunk BC, Ginsburg PB, Gabel JR. Tracking Health Care Costs. Health Affairs, 26 September 2001 (Web exclusive).

Synder TE, Selanders JR. Incidental appendectomy—yes or no? A retrospective case study and review of the literature. Infect Dis Obstet Gynecol 1998;6(1)30-7.

Taylor CL, Yuan A, Selman WR, et al. Mortality rates, hospital length of stay, and the cost of treating subarachnoid hemorrhage in older patients: institutional and geographical differences. J Neurosurg 1997;86(4):583-8.

The Center for Medical Consumers. (http://www.medicalconsumers.org/)

Tissue plasminogen activator for acute ischemic stroke. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. N Engl J Med 1995;333(24):1581-7.

Warren JL, Penberthy LT, Addiss DG, et al. Appendectomy incidental to cholecystectomy among elderly Medicare beneficiaries. Surg Gynecol Obstet 1993;177(3):288-94.

Wen SW, Simunovic M, Williams JI, et al. Hospital volume, calendar age, and short term outcomes in patients undergoing repair of abdominal aortic aneurysm: the Ontario experience, 1988-92. J Epidemiol Community Health 1996;50(2):207-13.

Whisnant JP, Sacco SE, O'Fallon WM, et al. Referral bias in aneurysmal subarachnoid hemorrhage. J Neurosurg 1993;78(5):726-32.

Williams GR, Jiang JG, Matchar DB, et al. Incidence and occurrence of total (first-ever and recurrent) stroke. Stroke 1999;30(12):2523-8.

Wolff BG. Current status of incidental surgery. Dis Colon Rectum 1995;38(4):435-41.

Wolinsky FD, Fitzgerald JF, Stump TE. The effect of hip fracture on mortality, hospitalization, and functional status: a prospective study. Am J Public Health 1997;87(3):398-403.

Ziskind AA, Lauer MA, Bishop G, et al. Assessing the appropriateness of coronary revascularization: the University of Maryland Revascularization Appropriateness Score (RAS) and its comparison to RAND expert panel ratings and American College of Cardiology/American Heart Association guidelines with regard to assigned appropriateness rating and ability to predict outcome. Clin Cardiol 1999;22(2):67-76.

Appendix A: Links

The following links may be helpful to users of the AHRQ Inpatient Quality Indicators.

Inpatient Quality Indicators Version 3.0 Documents and Software

Available at http://www.qualityindicators.ahrq.gov/iqi_download.htm

Title	***Description***
Guide to Inpatient Quality Indicators	Describes how the IQIs were developed and provides detailed evidence for each indicator.
Inpatient Quality Indicators Technical Specifications	Provides detailed definitions of each IQI, including all ICD-9-CM and DRG codes that are included in or excluded from the numerator and denominator. Note that exclusions from the denominator are automatically applied to the numerator.
IQI Covariates used in Risk Adjustment	Tables for each IQI provide the stratification and coefficients used to calculate the risk-adjusted rate for each strata.
SAS® IQI Software Documentation	This software documentation provides detailed instructions on how to use the SAS ® version of the IQI software including data preparation, calculation of the IQI rates, and interpretation of output.
SPSS® IQI Software Documentation	This software documentation provides detailed instructions on how to use the SPSS® version of the IQI software including data preparation, calculation of the IQI rates, and interpretation of output.
Change Log to IQI Documents and Software	The Change Log document provides a cumulative summary of all changes to the IQI software, software documentation, and other documents made since the release of version 2.1 of the software in March 2003. Changes to indicator specifications that were not a result of new ICD-9-CM and DRG codes, are also described in the Change Log.
Fiscal year 2006 Coding Changes	This document summarizes the changes to the indicator definitions resulting from FY 2006 changes to ICD-9-CM coding and DRG changes. These changes will only affect data from FY 2006 (October 1, 2005) or later.
SAS® IQI Software	Requires the SAS® statistical program distributed by the SAS Institute, Inc. The company may be contacted directly regarding the licensing of its products: http://www.sas.com
3M® APR® DRG Limited License Grouper for SAS®	Creates APR-DRG variables for use with SAS version of IQI software. Instructions for running the software are included in the Zip file.
SPSS® IQI Software	Requires the SPSS® statistical program distributed by SPSS, Inc. The company may be contacted directly regarding the licensing of its products: http://www.spss.com

AHRQ QI Windows Application

The AHRQ QI Windows Application calculates rates for all of the AHRQ Quality Indicators modules and does not require either SAS® or SPSS®. It is available at:

http://www.qualityindicators.ahrq.gov/winqi_download.htm

Additional Documents

The following documents are available within the "Documentation" section of the **AHRQ QI Downloads** Web page:

http://www.qualityindicators.ahrq.gov/downloads.htm

- *Refinement of the HCUP Quality Indicators (Technical Review), May 2001*
- *Refinement of the HCUP Quality Indicators (Summary), May 2001*
- *Measures of Patient Safety Based on Hospital Administrative Data - The Patient Safety Indicators, August 2002*
- *Measures of Patient Safety Based on Hospital Administrative Data - The Patient Safety Indicators (Summary), August 2002*

In addition, these documents may be accessed at the AHRQ QI Documentation Web page:

http://www.qualityindicators.ahrq.gov/documentation.htm

- *Guidance for Using the AHRQ Quality Indicators for Hospital-level Public Reporting or Payment, August 2004*
- *AHRQ Summary Statement on Comparative Hospital Public Reporting, December 2005*
- *Appendix A: Current Uses of AHRQ Quality Indicators and Considerations for Hospital-level*
- *Comparison of Recommended Evaluation Criteria in Five E*xisting National Frameworks

The following documents can be viewed or downloaded from the page:

http://www.qualityindicators.ahrq.gov/newsletter.htm

- *2006 Area Level Indicator Changes*
- *Considerations in Public Reporting for the AHRQ QIs*
- *June 2005 Newsletter* - Contains the article, "Using Different Types of QI Rates"

Other Tools and Information

IQI rates can be calculated using the modified Federal Information Processing Standards (FIPS) State/county code. A list of codes is available at:

http://www.census.gov/popest/geographic/codes02.pdf

AHRQ provides a free, on-line query system based on HCUP data that provides access to health statistics and information on hospital stays at the national, regional, and State level. It is available at:

http://hcup.ahrq.gov/HCUPnet.asp

Information on the 3M™ APR-DRG system is available at:

http://www.3m.com/us/healthcare/his/products/coding/refined_drg.jhtml