*Final Contract Report*

---

# Selecting Quality and Resource Use Measures: A Decision Guide for Community Quality Collaboratives

**Prepared for:**
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No./Assignment No. 08003967

**Prepared by:**

Patrick Romano, M.D., M.P.H.
*University of California, Davis*

Peter Hussey, Ph.D.
*RAND Corporation*

Dominique Ritley, M.P.H.
*University of California, Davis*

AHRQ Publication No. 09(10)-0073
May 2010

# Acknowledgments

This Decision Guide is intended for use by community quality collaboratives interested in evaluating quality and resource use measures. The guide presents answers to 26 questions, identified in collaboration with representatives from the Agency for Healthcare Research and Quality (AHRQ) Chartered Value Exchanges.

The authors gratefully acknowledge the following people for reviewing a formative draft of this document:

- Catherine Baase, Michigan Health Information Alliance
- Craig Brammer, Health Improvement Collaborative of Greater Cincinnati
- Karen DeSalvo, Louisiana Health Care Quality Forum
- R. Adams Dudley, University of California-San Francisco
- Rob Ence, Utah Partnership for Value-Driven Health Care
- David Kelley, Pittsburgh Regional Health Initiative
- Mark Legnini, Engelberg Center for Health Care Reform/The Brookings Institution
- Cathie Markow, California Chartered Value Exchange
- Aucha Prachanronarong, Center for Medicare & Medicaid Services
- Joachim Roski, Engelberg Center for Health Care Reform/The Brookings Institution
- Karen Shore, Center for Health Improvement
- David Shute, Oregon Health Care Quality Corporation
- Mark Sonneborn, Minnesota Healthcare Value Exchange
- Amy Topel, Wisconsin Healthcare Value Exchange

The following AHRQ staff members reviewed the document: Roxanne Andrews, John Bott, Christine Crofton, Katherine Crosson, Jan De La Mare, Anne Elixhauser, Irene Fraser, Peggy McNamara, Mamatha Pancholi, Jennifer Schnaier, Carol Stocks, and Herbert Wong.

William Encinosa, AHRQ, Michael Painter, Robert Wood Johnson Foundation, Nancy Brands Ward, Center for Health Improvement, and Nancy Wilson, AHRQ, provided targeted feedback. In addition, the authors extend their appreciation to Colleen Cameron, David Chin, and Tammy Gee, University of California, Davis; and Michele Peterson at the Center for Health Improvement for their valuable technical assistance.

---

Romano PS, Hussey P, Ritley D. Selecting Quality and Resource Use Measures: A Decision Guide for Community Quality Collaboratives. Rockville, MD: Agency for Healthcare Research and Quality; May 2010. AHRQ Publication No. 09(10)-0073.

# Contents

## Tables

## Figures

# Foreword: Improving Quality One Community at a Time

The Agency for Healthcare Research and Quality (AHRQ) partnered with Patrick Romano, M.D., M.P.H., Peter Hussey, Ph.D., and Dominique Ritley, M.P.H., to develop *Selecting Quality and Resource Use Measures: A Decision Guide for Community Quality Collaboratives*. Our collective motivation was to meet the needs of local health care leaders seeking an evidence-based primer and decisionmaking framework to guide their strategic and operational planning related to performance measurement.

The Decision Guide tapped a panel of 10 community-based leaders—representing purchasers, plans, providers, and consumer organizations—over the yearlong development period. The panel was asked to identify questions that need to be addressed when considering or designing a measurement strategy, which were then used to form the outline for the Decision Guide. Responses to these questions, developed by the expert team led by Patrick Romano, summarize empirical evidence, when it exists, and incorporate expert advice, best practices, and real life case examples to illustrate the breadth of considerations and implementation options.

The resulting Decision Guide is organized into five sections:

- Introduction to performance data
- Introduction to measures of quality
- Introduction to resource use/efficiency measures
- Selecting quality and resource use measures
- Interpreting quality and resource use measures

Partnering with local leadership to improve quality and value is not new to AHRQ. AHRQ actively supports community-based quality collaboratives through a portfolio of initiatives, the centerpiece of which is a Learning Network for Chartered Value Exchanges (CVEs). The 24 community quality collaboratives that participate in the Learning Network include active participation from four key stakeholder groups—providers, private and public purchasers, health plans, and consumer organizations—in pursuit of a communitywide system of health care performance measurement, transparency, and improvement. CVEs are involved in a variety of different strategies and approaches to improving quality. But the measurement of quality and resource use, the focus of this Decision Guide, is the keystone activity that undergirds all others and is common to all the participating quality collaboratives.

The 24 community quality collaboratives that are working with AHRQ through the Learning Network provide a window into the broader pool of community collaboratives. Contrary to what might be hypothesized, collaborative implementation does not appear to be constrained to any particular setting or market condition. The 24 collaboratives illustrate a breadth of contexts in which quality collaboratives are being formed across the country. Some of the communities that host collaboratives have one or two dominant health plans, while others are in more competitive markets. Some are urban, while others include a large rural component.[1]

The 24 collaboratives also illustrate a variety of approaches to both operational policy and strategy. Most collaboratives govern by consensus, although they vary in terms of how often they meet.

Collaboratives' sources of operating revenue include dues from local members, sale of collaborative products (e.g., data) or services, and grant funding from foundations, governments, and local stakeholders. Most collaboratives rely at least in part on in-kind resources from their stakeholder members.[2] Of the four categories of collaborative membership, consumer organizations tend to have the fewest representatives involved in the process.

Approaches to collecting data vary across the collaboratives and quality measures include a rich mix of structural, process, and outcome indicators. Collaborative models share some common features that affect decisionmaking related to data and measure selection (e.g., decisionmaking by consensus, reliance on in-kind contributions). Other elements vary widely, such as size of annual operating budget, menu of quality measures, and use of health information technology. Some collaboratives have overcome significant challenges to quality measurement and reporting. Although not commonplace, a smaller subset of collaborative pioneers is moving beyond public reporting by developing provider or consumer incentives to reinforce their respective quality agendas. This breadth of environmental contexts and range of design models suggests that collaboratives are adaptable and feasible to implement nationwide, but budgetary, political, and other challenges can temper the pace of progress.[1]

In addition to the Learning Network for Chartered Value Exchanges, AHRQ's library of past and current quality implementation partnerships includes, for example:

1. The AHRQ Learning Network on Quality-based Purchasing, which provided a forum for employers and State Medicaid agencies to learn about pay-for-performance best practices from experts as well as each other.
2. The AHRQ Quality Indicators Learning Institute, which provided a forum for discussing and facilitating the use of the AHRQ Quality Indicators (QIs) in statewide and regional programs that report hospital quality measures to the public.
3. The AHRQ Medicaid Medical Directors Learning Network, which provides a venue for clinical leaders of State Medicaid programs to connect with other organizations interested in using evidence-based medicine to make policy decisions that affect Medicaid programs.

This Guide is the latest in a series of user-driven guides developed by AHRQ to distill and summarize evidence, expertise, and implementation considerations for an audience of local decisionmakers. Other guides are:

- *Pay for Performance: A Decision Guide for Purchasers*, an evidence summary organized around 20 questions that span 4 phases of purchaser decisionmaking related to provider incentives: contemplation, design, implementation, and evaluation.[3]
- *Consumer Financial Incentives: A Decision Guide for Purchasers*, an evidence summary organized around 21 questions that reviews the application of incentives to 5 types of consumer decisions: selecting a high-value provider, selecting a high-value health plan, deciding among treatment options, reducing health risks by seeking preventive care, and reducing health risks by decreasing or eliminating high-risk behavior.[4]

These and other evidence-based resources for community quality collaboratives can be ordered by e-mailing AHRQPubs@ahrq.hhs.gov, calling AHRQ's Publications Clearinghouse at 1-800-358-9295, or downloading from AHRQ's Web site at www.ahrq.gov/qual/value/localnetworks.htm.

While AHRQ has been working with scientists who develop the evidence base and decisionmakers in the field who apply the evidence base to guide their actions to make a positive impact on quality, we as a Nation still have a long way to go to achieve a 21st century health care system that serves all Americans well. Unfortunately, we know little about the long-term impact of various collaborative strategies on quality of care or on collaboratives' sustainability over time. More attention and resources are needed to build the evidence base about best practices in translating measurement into performance improvement (e.g., via public reporting and payment incentives) and to disseminate what we know to more decisionmakers who are in a position to act on the evidence. Much remains to be done to better understand how communities can most effectively engage purchasers, plans, providers, and consumer organizations in applying performance measures through reporting, pay-for-performance, consumer incentives, and HIT initiatives.

AHRQ expresses appreciation to the team of Patrick Romano, Peter Hussey, and Dominique Ritley and the interdisciplinary panel of reviewers that included community collaboratives, Federal Agency representatives, and other expert colleagues. In publishing this Decision Guide, we hope to contribute to and advance an ongoing local and national dialogue related to how community collaboratives and their component stakeholders can improve performance measurement and quality of care. We hope this Decision Guide informs their deliberations, and we welcome feedback.

The logic of aligning local health care leadership interests through a community collaborative to achieve a common quality improvement or value-enhancing agenda resonates. While national policies can support quality of care, real improvement requires hard work at the local level and is accomplished only one community at a time.

Peggy McNamara
Senior Fellow, AHRQ
E-mail: Peggy.McNamara@ahrq.hhs.gov

May 2010

## References

1. McNamara P, De La Mare J, Moss D. Local, multi-stakeholder quality collaboratives: what can the U.S. experience tell us about their feasibility as part of a national effort to improve health care quality? Rockville, MD: Agency for Healthcare Research and Quality; May 2009. Working paper.
2. McNamara, P. CVE sustainability—Overview. Presentation at AHRQ's Third National Meeting of Chartered Value Exchanges, Washington, DC, July 2009.
3. Dudley RA, Rosenthal MB. Pay for performance: a decision guide for purchasers. (Final Contract Report) Rockville, MD: Agency for Healthcare Research and Quality; April 2006. AHRQ Pub. No. 06-0047. Available at: www.ahrq.gov/qual/p4pguide.htm.
4. Dudley RA, Tseng CW, Bozic K, et al. Consumer financial incentives: a decision guide for purchasers (Final Contract Report). Rockville, MD: Agency for Healthcare Research and Quality; November 2007. AHRQ Pub. No. 07(08)-0059. Available at: www.ahrq.gov/qual/value/incentives.htm

## Selecting Quality and Resource Measures:
## A Decision Guide for Community Quality Collaboratives

### Part I: Introduction to Performance Data

✓ Question 1. What data, including both national and State sources, are readily available to collaboratives for performance measurement at the hospital and physician levels?

✓ Question 2. What are the strengths and weaknesses of using administrative data, medical record data, and hybrid data?

✓ Question 3. What are the opportunities and challenges in building a multipayer/ multi-data source database or data warehouse?

✓ Question 4. Should a vendor be used for data collection and management? If so, what are the criteria for selecting a vendor?

✓ Question 5. How should a data auditing program be designed to ensure data quality?

✓ Question 6. How do HIPAA and other privacy regulations affect data collection and public reporting?

### Part II: Introduction to Measures of Quality

✓ Question 7. How are quality performance measures constructed, and what are the implications of how their numerators and denominators are specified?

✓ Question 8. What specific measures can be used to calculate physician performance at the individual or organization level?

✓ Question 9. What specific measures can be used to calculate hospital performance regionally or nationally?

✓ Question 10. What is the role and value of composite measures, and what are the most common approaches to constructing composites?

✓ Question 11. What is “risk adjustment” and how is it best applied?

✓ Question 12. What are the opportunities and challenges to using patient experience surveys to measure hospital or physician performance at the regional or State level?

✓ Question 13. What is the “Better Quality Information” pilot project, sponsored by the Centers for Medicare & Medicaid Services, and what can be learned from it?

### Part III: Introduction to Resource Use/Efficiency Measures

✓ Question 14. What are the main types of resource use measures?

✓ Question 15. What types of data are used to construct resource use measures? How is “cost” measured?

✓ Question 16. What is known about the validity of available resource use measures, including their advantages and disadvantages?

✓ Question 17. Which national groups are developing or endorsing resource use measures?

✓ Question 18. How have resource use measures been used to compare providers to benchmarks?

**Selecting Quality and Resource Measures:**
**A *Decision* Guide for Community Quality Collaboratives (Continued)**

## Part IV: Selecting Quality and Resource Use Measures

- ✓ Question 19. What national initiatives and forces are driving the standardization of quality measurement?
- ✓ Question 20. How can the Institute of Medicine’s six “quality domains,” the National Priorities Partnership’s six National Priorities, and Donabedian’s “structure, process, and outcome” typology be used to select appropriate measures of quality?
- ✓ Question 21. What are the roles and responsibilities of the organizations that endorse or approve measures versus those organizations that develop measures?
- ✓ Question 22. What criteria should we use when screening measures of quality for public reporting or other purposes?
- ✓ Question 23. Against which benchmarks should we measure our local performance?
- ✓ Question 24. When and how should providers review data before public reports are released?
- ✓ Question 25. What are the critical success factors for selecting useful performance measures?

## Part V: Interpreting Quality and Resource Use Measures

- ✓ Question 26. How can quality and resource use measures be evaluated together to help identify high-value and low-value providers?

# Introduction

Selecting quality of care and resource use measures is an important and challenging task for organizations striving to improve the quality of health care in their communities. This Decision Guide is designed to inform readers about the most critical issues to consider when selecting and adopting such performance measures.

The topic of measure selection presumes that readers already have a basic understanding and comfort with many of the issues related to performance measurement. For those readers who need an introduction to performance measurement, see the box on the next page, “Why is health care performance measurement important?” for a brief summary and links to more introductory resources.

## Decision Guide Format and Methodology

This Decision Guide attempts to answer 26 questions that are frequently asked, in slightly different ways, by leaders and stakeholders in community quality collaboratives across the country. These questions were fine-tuned by a panel of 10 reviewers from Chartered Value Exchanges (CVEs), representative of various geographic regions and stakeholder organizations, who were selected by the Agency for Healthcare Research and Quality (AHRQ). A draft of this Decision Guide was reviewed by the same panel and many content experts and was revised based on their suggestions.

We organized the questions into five sections that may be read sequentially or to suit the user’s needs. Part I: Introduction to Performance Data presents information about sources of data, strengths and weaknesses of data, and data management strategies. Part II: Introduction to Measures of Quality discusses issues focused on quality measure construction, risk-adjustment, and specific measures used to calculate physician and hospital performance. Part III: Introduction to Resource Use/Efficiency Measures presents information about the primary types of resource use measures and the national groups developing such measures. It also discusses how these measures are used to compare providers to benchmarks. Part IV: Selecting Quality and Resource Use Measures offers different typologies and criteria for community collaboratives to consider when selecting their measures. The guide concludes with Part V: Interpreting Quality and Resource Use Measures, which addresses ways the two measure types can be evaluated together to identify high-value providers.

## Parameters

We recognize that each collaborative must adapt its programs to local market conditions and local concerns about health care quality. We also recognize that collaboratives, including CVEs, are in different stages of evolution, depending on when and how they were established. Therefore, some of the 26 questions may be more relevant to some collaboratives than to others. We hope that this Decision Guide fosters additional discussion within and across communities, so that leaders and stakeholders in collaboratives can share their experiences and improve their programs.

This Decision Guide uses the Institute of Medicine’s (IOM) definition of quality as “the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge."[1] Based on this definition, **quality problems are broadly categorized as problems of overuse, underuse, and misuse**.[2] We also build on the IOM's six domains of high-performance health care for the 21st century: **safety, effectiveness, patient centeredness, timeliness, efficiency, and equitability**. These concepts and their applicability to selecting quality of care and resource use measures will be further explained in the sections that follow.

This Decision Guide focuses on **hospital and physician data and measures**, but some of the underlying principles or criteria may also apply to selecting health plan, nursing home, or home health quality measures. The sources of data and measures are numerous and constantly changing; Web links are embedded throughout the Guide with the intent of producing a "living" document that is as current as possible.

Before selecting measures, **stakeholders should come to agreement as to *why* they are measuring quality** because the goals of measurement can affect the types of measures selected and how they are prioritized. For example, programs focused on driving quality improvement within health care organizations may emphasize process-of-care measures, whereas programs focused on engaging consumers may emphasize outcome and patient experience measures, which are typically more salient to consumers.

### Why is health care performance measurement important?

Two interrelated factors justify the efforts of community quality collaboratives to assess health care performance: substantial variation in the quality of care and the cost of that care. Numerous studies have shown that unexplained variation in quality due to underutilization, overutilization, and inappropriate care leads to unnecessarily high mortality and morbidity rates.[3-6] In addition, the cost of health care in the United States exceeded $2.2 trillion in 2007, representing more than 16% of the gross domestic product.[7] One can surmise that a significant portion of these expenditures was misspent, as research has shown that high-quality care is often associated with lower or equal cost.[8, 9]

Over the last 20 years, performance measurement has gained momentum through the efforts of public and private sector stakeholders that are interested in improving care. Measurement results are used privately to guide quality improvement within organizations and publicly for policy planning. Within the last decade, public reporting and financial incentive (pay-for-performance) programs have emerged. Measurement supports the transparency required for accountability and quality improvement in the health care system. Efforts to measure quality and cost have thus become widespread throughout the public and private sectors.

Several sources for introductory information about performance measurement and quality of care can be found at:

**AHRQ**
www.ahrq.gov/qual/value/localnetworks.htm
www.ahrq.gov/qual/qualix.htm
www.ahrq.gov/qual/measurix.htm
https://www.talkingquality.ahrq.gov

**Commonwealth Fund**
www.commonwealthfund.org/Publications.aspx

**National Quality Forum**
www.qualityforum.org/

# Part I. Introduction to Performance Data

Health care performance data can be obtained from multiple sources, including State and Federal governments, national accrediting bodies, research organizations, professional associations, health plans, employers, vendors that pool data from multiple plans or employers, and directly from providers. The answers to Questions 1-6 provide a framework for evaluating and selecting existing data from these various sources. Our emphasis is on data that are ready to present as accepted performance measures or can readily be converted into such measures, rather than on data that could be used to develop entirely new measures. These performance data can be delivered as ready-to-use, provider-specific results or as raw data requiring data management and analytic expertise to generate such results.

## Question 1. What data, including both national and State sources, are readily available to collaboratives for performance measurement at the hospital and physician levels?

Within the last several years, key organizations have led a **national public-private effort to harmonize measures** to reduce the data collection burden on providers and to minimize confusion among stakeholders.[10,11] Thus, many data reporting organizations overlap in the measures they collect and the results they provide. This answer summarizes the variety of available data sources for measuring hospital and physician performance. (Refer to the responses to Questions 8 and 9 for related information about measures.)

Most public agencies that offer off-the-shelf, summarized data at the provider level use measurement and data collection processes that have undergone rigorous review by committees of stakeholders. Using these existing sources obviates the need for a collaborative to undertake the time-consuming and costly task of validation. In addition, the data from these sources are frequently free. However, such data sources have several disadvantages, including inadequate coverage of some performance domains (domains described further in response to Question 20) and "one-size-fits-all" presentation of data, which may not reflect local or regional priorities.

To fill these gaps, some private entities have developed and implemented their own data collection and analysis systems, as further described in Table 1. However, the validity of these systems has generally not been established, and voluntary provider participation is a major limitation. Lacking any mechanism for auditing or any penalties for incorrect reporting, these systems are likely to suffer from selective nonparticipation by poorly performing providers and selective overreporting of performance.

### *Hospital Data*

Table 1 summarizes off-the-shelf, publicly available sources of data available for hospital performance measurement at the national or State level. Some are available for free from government agencies, whereas others may require contracts with proprietary organizations.

**Table 1. Organizations providing hospital-level data**

<table>
<tr><th colspan="2">National Data Sources</th></tr>
<tr><td><b>Centers for Medicare & Medicaid (CMS): Hospital Compare</b></td><td><b>www.hospitalcompare.hhs.gov</b></td></tr>
<tr><td colspan="2">CMS, in collaboration with the Hospital Quality Alliance, supports the HospitalCompare Web site, which reports hospital-specific data for clinical process-of-care indicators (e.g., most of The Joint Commission’s “Core Measures”), clinical outcomes indicators (e.g., risk-adjusted mortality and readmission rates for selected conditions), and patient experience measures derived from the HCAHPS® (Consumer Assessment of Healthcare Providers and Systems Hospital Survey). More information on the CAHPS patient experience survey can be found at www.hcahpsonline.org. In addition, CMS piloted the reporting of selected AHRQ QIs in 2009. All of these efforts have been undertaken under the auspices of a program known as Reporting Hospital Quality Data for Annual Payment Update (RHQDAPU), which links annual market basket payment updates to hospital participation.<br><br><b>Comment:</b> Much of the process-of-care data available on the Hospital Compare Web site is also available from The Joint Commission. The outcomes data are limited in distinguishing provider quality, because they are based only on Medicare fee-for-service beneficiaries who received inpatient care at nonfederal hospitals.</td></tr>
<tr><td><b>The Commonwealth Fund: “Why Not the Best?”</b></td><td><b>www.whynotthebest.org</b></td></tr>
<tr><td colspan="2">Announced in 2008, The Commonwealth Fund's hospital performance reporting Web site repackages the information provided through Hospital Compare and offers hospital-specific composite scores for heart attack care, pneumonia care, heart failure care, surgical care improvement, and patient experience, and an overall performance score.<br><br><b>Comment:</b> For collaboratives just starting to consider public reporting, the Commonwealth Fund’s “opportunity-weighted” composites offer a relatively easy-to-understand approach for presenting complex data on processes of care.</td></tr>
<tr><td><b>HealthGrades®</b></td><td><b>www.healthgrades.com</b></td></tr>
<tr><td colspan="2">HealthGrades® is perhaps the best known comparative reporting system. It assigns star ratings of “best,” “as expected,” or “poor” for at least 29 procedures and diagnoses using HealthGrades Hospital Report Card™ Mortality and Complication Based Outcomes Methodology. Most of these analyses are performed by applying proprietary risk-adjustment models to publicly available Medicare claims or all-payer hospital discharge data. HealthGrades also confers its Distinguished Hospital Awards for superior performance on a composite of selected AHRQ Patient Safety Indicators, Specialty Excellence Awards for superior performance on relevant risk-adjusted mortality or complication measures, and Outstanding Patient Experience Awards for superior performance on a composite of HCAHPS® items.<br><br><b>Comment:</b> HealthGrades offers very limited information for public use through its Web site, although more detailed information may be available for purchase. The limited available information about HealthGrades’ methods is a potential concern, and most measures are based only on Medicare fee-for-service claims.</td></tr>
</table>

**The Joint Commission: Quality Check™** **www.qualitycheck.org**

The Joint Commission is a not-for-profit hospital accrediting body that provides data on its “Core Measures,” most of which focus on specific evidence-based processes of care, as well as HCAHPS® survey data. Hospitals are required to submit data quarterly and to meet other specified standards to receive accreditation and certification. Quality data can be downloaded directly through the Quality Check Web site.

**Comment:** There is overlap between measures reported by The Joint Commission's Quality Check and CMS's Hospital Compare Web sites. Because the Core Measures are divided into seven sets, and hospitals are currently only required to report four, not all measures are available from all hospitals.

**Leapfrog Group** **www.leapfroggroup.org/cp**

Leapfrog conducts a voluntary annual survey of hospitals nationwide to assess performance based on four quality and safety practice domains that are believed to reduce preventable medical mistakes. Endorsed by the National Quality Forum (NQF), these practice domains include computerized physician order entry, intensive care unit staffing by “intensivist” physicians, evidence-based hospital referral for selected high-risk procedures, and a targeted subset of 34 “safe practices” endorsed by the NQF. The resulting data are currently summarized for 10 different conditions and procedures, two adverse outcomes (hospital-acquired pressure ulcers and fall-related injuries), and a composite measure of “steps to avoid harm.” In addition, Leapfrog assigns a rating of hospitals’ implementation of its Policy Statement on Serious Reportable Events/“Never Events” (which identifies events that should never happen [e.g., removing the wrong limb or leaving surgical equipment inside a patient after surgery]).

**Comment:** The Leapfrog survey is conducted by Thomson Healthcare. Voluntary responses were submitted in 2008 from 1,276 hospitals in 37 major U.S. metropolitan areas, representing 48% of urban, general acute-care hospitals. Most components of the survey are not audited, so reporting bias is a potential concern.

**Thomson Reuters** **www.100tophospitals.com/**

Thomson Reuters (formerly Solucient) recognizes “100 Top Hospitals” based on both risk-adjusted clinical outcomes and financial performance, using publicly available Medicare claims or all-payer hospital discharge data. In addition, Thomson Reuters and other health care consultants contract with large national employers and their coalitions to analyze commercial claims data. These data may be particularly useful in midsize communities dominated by some of these large national employers.

**Comment:** Very limited information for public use is available through this Web site, although more detailed information may be available for purchase. The limited available information about Thomson’s methods is a potential concern.

**U.S. News & World Report** **http://health.usnews.com/sections/health/best-hospitals/**

U.S. News & World Report annually evaluates the performance of more than 1,500 U.S hospitals in at least 16 specialties. In four of these specialties, evaluations are based purely on reputation, as reported by a national sample of physicians. In the other 12, evaluations are also based on risk-adjusted mortality from Medicare claims data (estimated using 3M® Health Information Systems APR-DRG* software), some AHRQ QIs, and other care-related factors, such as volume and nurse staffing (ascertained using the American Hospital Association's annual survey of hospitals).

**Comment:** Very limited information for public use is available through this Web site, although more detailed information may be available for purchase. The limited available information about U.S. News’ methods is a potential concern.

*All Payer Refined-Diagnosis Related Group

## State or Regional Data Sources

### Statewide Health Data Organizations

Sources of hospital data at the State level include statewide health data organizations, hospital associations, private data organizations (e.g., Hawaii and Virginia), and State departments of health. These entities often collect hospital utilization, financial, and structural data relevant to hospital performance, in addition to all-payer inpatient hospital discharge data that go into AHRQ's multistate Healthcare Cost and Utilization Project (HCUP) databases. Some States (e.g., Pennsylvania) are collecting data on hospital-acquired infections or serious reportable events, while others are collecting data on nurse staffing levels or nursing skill mix. Most of these data collection efforts incorporate definitions established by the Centers for Disease Control and Prevention or the National Quality Forum, but the definitions may still vary slightly by State.

Inviting representatives from the State's health data organization to partner with the community quality collaboratives may yield useful data for local or regional analysis that would be unavailable through a national source. A relatively low-cost approach to analyzing State-level data is to use AHRQ's Inpatient Quality Indicators, Pediatric Quality Indicators, and Patient Safety Indicators (refer to Question 9), because Windows-compatible software and documentation can be downloaded for free from the Quality Indicators Web site (http://qualityindicators.ahrq.gov). In 2010, AHRQ will release a downloadable program (MONAHRQ) that allows organizations to input their hospital discharge data and output a Web site that provides query-driven information on all AHRQ QIs, including potential safety-related events, preventable hospitalizations, and volume or utilization of specific hospital services.

**Comment:** Rules regarding the release of hospital-specific data differ across States. Community quality collaboratives need to contact their statewide health data organization (www.hcup-us.ahrq.gov/partners.jsp) to determine such availability. Other limitations of this data source are that the data may not be available on a timely basis, depending on the State, and the quality of diagnosis coding for adverse events may vary across hospitals (which is particularly relevant to the AHRQ Patient Safety Indicators). However, these data offer the important advantage of being population based and including all payers (not just Medicare or Medicaid).

### State and Regional Coalitions

Some States and regional coalitions have developed their own mechanisms for repackaging and rescoring hospital-specific data from Hospital Compare and from The Joint Commission. For example, the California Hospital Assessment and Reporting Task Force (CHART: www.calhospitalcompare.org) adds State-level all-payer data on risk-adjusted coronary artery bypass surgery, adult critical care, and inpatient pneumonia mortality, along with rates of breastfeeding without formula supplementation (based on requests for genetic disease screening), and then rescores hospital performance as superior, above average, average, below average, or poor. Different cut points and benchmarks can be used by different reporting organizations, resulting in assignment of different symbols or ratings (stars, checkmarks, etc.), even though the underlying numeric scores are equal.

**Comment:** If a community quality collaborative chooses to draw upon multiple data sources in this manner, the scoring methodologies used by the reporting organizations should be carefully evaluated. Using the same cut points or benchmarks as the original data sources may lead to internal inconsistencies in how performance data are presented, but changing cut points or benchmarks may lead to inconsistencies with other sites reporting the same indicators on the same hospitals.

**Proprietary Data (Providers/Business Coalitions/Health Plans/Consultants)**

Proprietary data from health plans, providers, and actuarial and health care decision support firms may be available to community quality collaboratives. Inviting these types of organizations to work with the collaborative may provide access to unique data and data management expertise. Successful partnerships with these providers have been established by Chartered Value Exchanges (CVEs) in Wisconsin, Oregon, and California. In addition, some CVEs collaborate closely with local employer coalitions to assemble rich data sets that capture a significant portion of the commercially insured population. For example, Wisconsin’s The Alliance has produced performance data in this manner for years, with resulting improvements in targeted domains of health care quality.

**Comment:** If a community quality collaborative has not already done so, we recommend inviting these organizations as partners in performance measurement.

### *Physician Data*

As Table 2 illustrates, national efforts to collect physician group or individual physician-level quality data are in the experimental stage due in part to technical issues with limited sample sizes and standardization of IT systems. National and regional initiatives are underway to enhance the usefulness of existing data for physician performance measurement. For example, the Physician Quality Reporting Initiative (PQRI) (www.cms.hhs.gov/PQRI) offers incentive payments to physicians who report at least some quality measures (from a list of approximately 179 in fiscal year 2010) applicable to Medicare fee-for-service beneficiaries. But only 16% of eligible professionals participated in 2007, and only 52% of those who participated met the program and reporting requirements. Data results are not yet publicly available, and provider participation will need to improve before CVEs will find the data useful. In particular, the very limited number of measures on which any individual physician must report is a major limitation for comparative reporting.

**Table 2. Organizations providing physician-level data**

| National Data Sources | |
|---|---|
| **Generating Medicare Physician Quality Performance Measurement Results** | **www.cms.hhs.gov/GEM** |
| As part of a project called "Generating Medicare Physician Quality Performance Measurement Results" (GEM), the Centers for Medicare & Medicaid Services (CMS) contracted with Masspro to calculate medical group practice performance results based on 2006-2007 Part B claims for fee-for-service Medicare beneficiaries. Information is presented at the population level, by State and ZIP code, and at the national level. These data are publicly available and may be downloaded from http://www.cms.hhs.gov/GEM/ . Results were calculated for the following measures using the National Committee for Quality Assurance’s (NCQA) Healthcare Effectiveness Data and Information Set (HEDIS) definitions of these measures: Breast Cancer Screening, LDL Testing for Diabetics, Retinal Eye Exam for Diabetics, HbA1c Testing for Diabetics, Cardiovascular LDL Testing, Colorectal Cancer Screening, Nephropathy Testing for Diabetics, Persistence of Beta Blocker Therapy- Post MI, Annual Monitoring for Patients on Persistent Medications, Anti-Depressant Medication Management-Acute Phase, Beta Blocker Treatment After Heart Attack, and Disease-Modifying Anti-Rheumatic Drug Therapy. | |

**Comment:** The data formats were designed to allow performance results at the group practice level (i.e., Taxpayer Identification Number) to be aggregated with similar data from commercial sources, but participating CVEs encountered difficulties with provider attribution. The future of this effort is unclear, because the CMS unit of reporting, Taxpayer Identification Number, does not correspond to individual physicians, physician practice sites, physician organizations, or any other unit recognizable to consumers.

| **Consumers' Checkbook** | **www.checkbook.org/doctors/pageone.cfm** |
|---|---|

Given the relative dearth of publicly available data on physician performance from official sources, other organizations are attempting to meet the market need. Consumers' Checkbook asks "roughly 260,000 physicians to tell us which specialists they would want to care for a loved one." They use the survey responses to construct a "Top Doctors" database, which contains the names of more than 20,000 doctors who were mentioned most often, across 30 specialties and 50 metropolitan areas. Consumers' Checkbook is also piloting an abbreviated version of the Clinician/Group (C/G) CAHPS® tool in three sites, although the reliability and validity of this abbreviated tool are unknown.

**Comment:** The reliability and validity of peer assessments of physician performance are not well established, given the poor response rate for most physician surveys.

| **HealthGrades®** | **http://www.healthgrades.com** |
|---|---|

HealthGrades offers an interactive survey tool (based loosely on C/G CAHPS®) for users to describe their experiences with individual physicians and whether they would recommend the physician to family or friends. These data are offered along with information about physicians' board certification (also available from the American Board of Medical Specialties at http://www.abms.org/WC/Login.aspx), group practice and hospital affiliations, insurance plans accepted, and licensure and disciplinary actions by State medical boards (also available from the Federation of State Medical Boards or http://www.docboard.org/docfinder.html).

| **National Committee on Quality Assurance Physician Recognition Program** | **www.ncqa.org/** |
|---|---|

The National Committee on Quality Assurance (NCQA) Physician Recognition Program publicly recognizes individual physicians who meet clinical requirements for appropriate care in back pain; heart/stroke; and diabetes; and who establish a primary care medical home. In addition, NCQA also offers a Physician and Hospital Quality certification program to health plans that evaluate the cost and quality of physicians and hospitals. The list of certified health plans that evaluate physician care can be found at www.ncqa.org/tabid/954/Default.aspx (login required).

**Comment:** NCQA's reporting tool, QualityCompass, does not report HEDIS measures for physicians. However, some local collaboratives have adapted the measures for physician performance measurement, most notably the Wisconsin and California CVEs.

| **Vitals.com** | **http://www.vitals.com** |
|---|---|

Vitals.com aims to present a 360° view of physicians, including factual information about their background, consumer reviews, peer reviews and awards, and office information. Vitals.com also offers an interactive survey tool with questions about the appointment process, waiting time, staff professionalism, accuracy, bedside manner, adequacy of time, and followup. Consumers are encouraged to write free-text reviews. Similar to HealthGrades, this site provides information about board certification, hospital affiliations, and insurance plans accepted, but it adds a publication list for physicians who have authored peer-reviewed papers.

**Comment:** HealthGrades' and Vitals.com's patient survey tools offer different response options than C/G CAHPS, even when the questions are similar, limiting one's ability to compare results across tools. Other proprietary sites, including Angie's List and Zagat, have more recently entered this market. For example, the managed care company WellPoint recently contracted with Zagat to encourage members of its "consumer-driven" Blue Cross plans in Los Angeles, Cincinnati, Dayton, and Connecticut to rate physicians on four distinct attributes: Trust, Communication, Availability, and Environment. Both Zagat and Angie's List encourage free-text ratings, although the latter site is available only to paid subscribers.

## State Data Sources

### Community Quality Collaboratives (e.g., Chartered Value Exchanges) and Local Health Care Coalitions

Across the country, community quality collaboratives and local health care coalitions are laying the foundation for collecting and reporting physician performance data. These organizations build their own network of local health plans, employers, and physician organizations to coordinate data collection and analysis. For example, the Washington-Puget Sound Health Alliance reports clinic and medical group performance using 21 indicators. California's Integrated Healthcare Association created a pay-for-performance system that integrates 12-13 HEDIS measures based on claims data, 4-5 measures of information technology-enabled "systemness," 6 new resource use measures, 9 patient experience measures, and 8-9 measures of coordinated diabetes care. Similarly, the Wisconsin Collaborative for Health Care Quality reports a mixture of HEDIS measures and its own measures. Some health plans use internal claims data to rate contracted physicians or physician organizations by applying HEDIS definitions. However, the availability and content of these ratings vary across plans and across States.

**Comment:** Community quality collaboratives are encouraged to partner with health plans and employers to build multipayer claims databases that can be used to evaluate process-of-care measures at the physician and physician organization levels. Such databases should be constructed to include as many payers as possible to improve the reliability of the resulting estimates and to reduce the possibility of conflicting data on the same physicians from different sources.

### State or Local Registries

A handful of States maintain disease or procedure registries, most commonly focused on coronary artery bypass graft (CABG) surgery and related "open heart" procedures, to enable public reporting of surgeon-specific risk-adjusted outcomes. These States include New York, California, and Pennsylvania. Other States could establish similar registries, building on data management infrastructures established by medical specialty organizations, such as the Society for Thoracic Surgeons and the American College of Cardiology.

**Comment:** Registries for CABG surgery have been in place for more than a decade, and the risk-adjustment models using these data have been repeatedly refined and validated. However, these registries are generally limited to narrowly defined subsets of patients, which may limit their utility.

### State medical boards

State medical boards generally maintain their own databases with licensure and disciplinary actions, including basic information submitted as part of the licensure process (e.g., medical school and year of graduation, residency training, and board certification). The format and structure of these data vary by State. Commercial Web sites, such as Vitals.com and HealthGrades.com, frequently list this information as part of the quality information they present. In addition, this information is used to populate the American Medical Association's DoctorFinder site (http://webapps.ama-assn.org/doctorfinder) and the Administrators in Medicine (Association of State Medical Board Executive Directors) DocFinder site (http://www.docboard.org/docfinder.html).

**Comment:** These sources provide very basic information about physicians' training and experience but do not directly address the quality of care that physicians provide.

## Question 2. What are the strengths and weaknesses of using administrative data, medical record data, and hybrid data?

Four (commonly used or primary) sources of data are used to assess health care quality, three of which are described here: administrative, medical record, and hybrid data.[12] (Survey data are also used to assess quality of care and will be addressed in Question 12 of this guide). In general, measure specifications should guide the selection of the most appropriate data format to ensure the most reliable and valid results.

### *Administrative Data*

**Administrative data are derived from a variety of preexisting sources** such as insurance enrollment files and provider claims. These data are:

- Readily available,
- Relatively inexpensive to acquire in electronic formats,
- Coded by health information professionals using accepted coding systems, and
- Drawn from large populations and therefore more representative of the populations of interest.[13]

Because most administrative data are intended for financial management rather than quality assessment, they contain varying degrees of clinical detail and are often limited in content, completeness, timeliness, and accuracy.[14] Studies analyzing the validity of administrative data for quality assessment at the hospital and health plan levels have generally found that administrative data are sufficiently sensitive and specific to estimate certain performance measures, such as mammography or prenatal care rates.[14]

**The quality of administrative data is likely to be better if the data originate from hospitals than if the data originate from physician offices or other ambulatory settings.** Hospitals employ professional coders to assign diagnosis and procedure codes, submit all-payer data in most States to health data agencies, and are subject to auditing and financial penalties for incorrect reporting.[15-17] For example, Steinwachs, et al., found that Medicaid administrative data undercounted visits by 25% for low-cost providers and by 41% for patients with low utilization, whereas medical records undercounted billed visits by 10% for patients with high utilization.[18] In other words, a significant proportion of ambulatory services may not generate a claim, especially in systems where provider payments are bundled or capped. These limitations of administrative data may be compounded when measuring performance at the individual physician level, as sample sizes are small and patient populations are heterogeneous.[19]

### *Medical Records Data*

**Obtaining data from medical records (paper or electronic) requires expert staff and greater financial and time resources.** Medical reviewers, who are typically either nurses or physicians, must interpret each record and input data findings. Medical records provide detailed clinical data with a richer description of care than can be obtained from administrative data. This is useful for some quality measures, especially those that rely on laboratory values (e.g., hemoglobin A1c [HbA1c] or cholesterol levels) and specific treatments (e.g., discharge instructions). The Core Measures of hospital performance, which were defined by The Joint Commission, endorsed by the National Quality Forum, and adopted by the Hospital Quality

Alliance and the Centers for Medicare & Medicaid Services (CMS), rely principally on medical record data. However, ICD-9-CM (International Classification of Diseases, Ninth Revision, Clinical Modification) coded administrative data are often used to help identify the denominator cases of interest. The cost of collecting and submitting these data is borne by hospitals as part of the accreditation process.

**Many hospitals and physician organizations voluntarily participate in clinical registries** that also capture abstracted data from medical records and report performance data confidentially to participating providers. Although these registry data are not generally used for public reporting or value-based purchasing, their use is under discussion in many regions. In some cases, for example, health plans or employer coalitions may designate centers of excellence for specific services based on participation in the appropriate registry and voluntary sharing of summarized outcome data. Such use of registries is still very limited, as it is generally discouraged by national registry sponsors. Examples of registries include the American College of Surgeons' National Surgical Quality Improvement Program and the Society of Thoracic Surgeons' Adult Cardiac Surgery and Congenital Heart Surgery Databases. The United Network for Organ Sharing collects and manages detailed clinical data pertaining to waiting lists and outcomes for organ transplantation, but only selected center-specific information is available to the general public.

### *Hybrid Data*

Both administrative and medical record data alone have limitations for measuring quality of care. **Hybrid data bring together both administrative data and medical record data to build on the strengths of each and to compensate for some of their respective weaknesses**.[18,20] Varying definitions exist to describe hybrid data, but the term typically refers to aggregation of electronic claims and information obtained from either electronic or paper medical records to increase the number of relevant data elements or to reduce the number of records that must be reviewed, the time required to review each record, or both.[14,21] Because organizations differ in their definitions of hybrid data, it is important to clarify expectations and procedures before pursuing this strategy.

At the physician level, applications of hybrid data generally involve **using claims to identify patients with a relevant diagnosis or problem and using medical records to identify specific clinical findings or nonpharmacologic treatments**. Relying on administrative data alone to estimate Healthcare Effectiveness Data and Information Set (HEDIS) indicators, Pawlson, et al., found significant underestimation and instability in health plan rankings, compared with results from hybrid data. For example, only 3 of 15 measures evaluated (all of which related to well child visits) had comparable performance estimates based on administrative and hybrid data. Evaluating diabetes care in the Veterans' Affairs system, Kerr, et al., compared administrative, medical record, and hybrid data at the Veterans Integrated Service Network (VISN) level and found high agreement between administrative and medical record data, but administrative data consistently underestimated facility performance.[22] Hybrid data yielded estimates similar to those from medical record data alone but required 50% fewer chart reviews, resulting in a significant cost reduction.[23]

At the hospital level, applications of hybrid data generally involve **combining ICD-9-CM coded administrative data with key laboratory or other clinical data to enhance the performance of risk-adjustment models** and to reduce bias in estimates of hospital performance, relative to administrative data alone.[24] The most helpful and cost-effective variables to collect for this purpose include blood cell counts, electrolytes, arterial blood gas values, clotting parameters, and vital signs.[25-27] An AHRQ-funded, multicenter pilot project involving Florida, Minnesota, and Virginia will enhance the utility of hybrid data for hospital-level analyses by: (1) standardizing collection of laboratory data using common nomenclature; (2) merging laboratory data with hospital administrative data; (3) assessing the added value of using clinical data to evaluate the quality of patient care within hospitals; and (4) developing recommendations for other sites. AHRQ will release a summary report of the experiences of the three participating organizations and a related toolkit in 2010.

### *The Future: Electronic Health Records*

Looking to the future, **electronic health record (EHR) systems may reduce the cost of accessing clinical information from the medical record, thereby making medical record data more useful for quality reporting**. However, there are hurdles to be overcome related to the interoperability of these systems and the continuing use of paper notes for point-of-care documentation in many hospitals and offices with EHR systems. EHR capabilities are still only partially implemented in most hospitals; for example, only 8% to 17% of hospitals have fully implemented computerized physician order entry, and fewer than 2% of hospitals have a comprehensive system present in all clinical units.[28,29] Other strategies to reduce the data collection burden for hospital measures that require medical record review have been proposed[21] but still need systematic evaluation.[30] In addition, some EHR data do not include information found in administrative data (e.g., charges and cost) that are important to public reporting, so there will remain a role for administrative data for the foreseeable future.

In ambulatory care, adoption of EHR systems in the United States has lagged behind the hospital sector (with only 4% of physicians reporting a fully functional system) and well behind other countries (e.g., the United Kingdom, Netherlands, Australia, and New Zealand have achieved more than 90% adoption among general practitioners).[31,32] However, the “meaningful use” provisions of the American Recovery and Reinvestment Act, which are tied to significant Federal incentive payments, promise to accelerate adoption over the next several years.

## Question 3. What are the opportunities and challenges in building a multipayer/multi-data source database or data warehouse?

Multisource databases offer the potential to analyze health outcomes and to estimate complex metrics that address the effectiveness and efficiency of care, rather than just specific steps in the process of care. Until now, most collaboratives have built such databases by aggregating data from multiple sources (i.e., employers or health plans) into what has been described as a “data warehouse.” A newer alternative under exploration is the distributed database, which could allow users to create a virtual database by pulling, in real time, only the required data for a particular query from disparate sources. Table 3 describes the differences between the conventional warehouse database and the distributed database.

In the near future, the infrastructure supporting data retrieval for both static, aggregated databases and distributed databases will be affected by the national effort to establish health information exchanges (HIEs).[33] The Federal Government is supporting a **Nationwide Health Information Network** (NHIN) in limited production. The NHIN will facilitate the exchange of health care information between State and regional HIEs, integrated delivery systems, health plans, personally controlled health records, Federal agencies, and others.[34] Once fully operational, the NHIN HIE specifications, testing materials, and trust agreements will be placed in the public domain to stimulate adoption (http://healthit.hhs.gov/portal/server.pt). Some areas of the United States have already seen successful information exchange through **Regional Health Information Organizations (RHIOs)**. RHIOs facilitate information sharing among enrolled members using common, nonproprietary standards for data content and exchange over existing networks and the Internet.[30] (Question 6 addresses privacy regulations, which also affect the data aggregation and sharing process.)

### *Conventional Aggregated Databases*

A conventional **aggregated database** is the most common structure used to warehouse and analyze health care performance data. Obtaining data from multiple sources (e.g., payers, providers, State databases) requires coordination between each source and the community collaborative (and frequently its vendor or consultant) that performs the analysis. This aggregated database approach has been successful in some settings, but it poses the hazard of inadvertently releasing protected health information during data transfers. Also, this approach can be associated with significant delays in obtaining timely data.

Examples of this approach include Indiana’s Quality First program, which is supported by its HIE and has been well received by users.[35] California’s RHIO recently published a white paper describing its success and sustainability model.[36] Additionally, many State public health departments that house surveillance databases are adopting this approach and collaborating with State RHIOs.[37]

### *Distributed Databases*

Issues such as financial responsibility, varied architectures, patient confidentiality, and data management responsibilities pose significant challenges to building and maintaining a multisource, aggregated database.[38] A **distributed database** operates as a “virtual” database where data from various sites (multiple health plans, physician offices, labs, etc.) remain onsite with the data owner. The data user can pull appropriate data from each of those sites, as needed in real time, and perform the necessary analyses from his or her desktop. Shared software, which must be installed by all participants, matches related data from key data sources while keeping protected health identifiers with the data owners, thereby decreasing the risk of disclosing sensitive information.[39]

State public health departments have used distributed databases to track infectious diseases, but only recently has this method been recognized as an approach for analyzing the performance of health care providers. The Quality Alliance Steering Committee and America’s Health Insurance Plans Foundation (and others) are testing a distributed database as part of their goal to develop a “nationally-consistent data aggregation methodology” that integrates data from multiple sources (www.healthqualityalliance.org/hvhc-project). One goal of this High-Value Health Care Project is to make performance information available as quickly, consistently, and efficiently as possible by determining the most accurate and timely sources of diagnostic and treatment information (e.g., registries, administrative data).

**Table 3: Comparison of multipayer database formats**

| | **Conventional (Aggregated) Method** | **Distributed Data** |
|---|---|---|
| **Examples** | Hospital quality data reporting programs, such as:<br>• California Hospital Assessment and Reporting Task Force (CHART)<br>• Pennsylvania Health Care Cost Containment Council | • Quality Alliance Steering Committee High-Value Health Care Project (www.healthqualityalliance.org/)<br>• CDC Vaccine Safety Datalink Project (www.cdc.gov/vaccinesafety/Activities/VSD.html)<br>• Florida public health surveillance program |
| **Data Collection** | Data owner collects and sends data to offsite location | Data users extract deidentified data that remain with the data owner. Data queries requiring patient identifiers occur within the data owner's domain; data users extract query results stripped of these identifiers. |
| **Data Location** | Data are physically transferred to offsite location | Data remain with each data owner; data for analysis reside with data user |
| **Advantages** | • Less expensive alternative in short run<br>• Familiar to users<br>• May be available in public use file<br>• Can accommodate new questions immediately<br>• Possible to repeat analyses at any time without concern that underlying data have changed | • Real-time access to data<br>• Reduces potential for HIPAA*/privacy violations<br>*HIPAA = Health Insurance Portability and Accountability Act |
| **Disadvantages** | • Data collection/ aggregation delays | • Startup costs may be significant; more complex or customized data use agreements may be needed<br>• Software updates to multiple users must be installed simultaneously, requiring coordination<br>• New uses or questions are likely to require new algorithms<br>• Data owners must agree not to delete or change data files or results will not be replicable |
| **Technology Requirements/ Architecture** | Software or system architecture must meet specification requirements determined by the data aggregator | Singular, master software program must be implemented concurrently at all participating data sites |

## Question 4. Should a vendor be used for data collection and management? If so, what are the criteria for selecting a vendor?

Many community quality collaboratives contract with vendors to assist with collecting and managing health care quality data. Vendors that have a solid understanding of the functional requirements and standards for capturing quality measurement data are a valuable resource for community quality collaboratives, especially those that plan to undertake their own data analyses and to coordinate regional measurement and public reporting.

### *Key Considerations for Vendor Selection[11,40]:*

1. **Issue a clear statement of the collaborative's goals and purpose** (e.g., pay-for-performance, public reporting, internal reporting, and quality improvement).
2. **Issue a clear statement of needs, expectations, and potential challenges** for the collaborative project ("request for proposals"), including:
   - Data collection procedures (e.g., claims data, hybrid data, surveys from multiple sources; frequency; volume; access to low-quality or high-quality claims data);
   - Data management procedures (e.g., data cleaning methods, data protection),
   - Data evaluation and validation procedures (e.g., use of simple or complex measure methodologies); and
   - Data storage (through vendor or in-house; appropriate security).

3. **Issue a clear statement of internal resources available to the collaborative** (e.g., level of in-house expertise).
4. **Ensure that the vendor has expertise** in collecting and managing quality performance data with established hardware and software systems (e.g., completed licensure or accreditation from quality measurement entities such as the National Committee on Quality Assurance, Quality Improvement Organizations, or the Joint Commission).
5. **Confirm vendor's data validation or auditing experience.**
6. **Confirm that no conflicts of interest exist** between collaborative members and the vendor, or within the vendor's various lines of business (e.g., vendor owning, owned by, or dependent upon a hospital or health plan that might be evaluated by the chartered value exchange [CVE]).
7. **Compare cost and services** offered by competing bidders. For government programs such as Medicaid, State rules may require a particularly complex process for soliciting and reviewing competitive bids.
8. **Request *and* contact references** provided by vendor, asking salient questions about the vendor's responsiveness, expertise, timeliness, quality, and financial management.
9. **In the case of Medicaid**, the vendor needs to understand the added complexities and challenges associated with Medicaid databases (e.g., issues of discontinuous eligibility, variable cost-sharing, nonstandard claims).

## Question 5. How should a data auditing program be designed to ensure data quality?

A carefully designed data auditing program will help to ensure the validity of the data reported and will proactively address concerns about data validity from both provider and consumer perspectives.

**Most national measurement efforts, such as those sponsored by the National Committee on Quality Assurance (NCQA) and the Joint Commission, involve a systematic data auditing program**. Two approaches have been developed. In the **decentralized approach**, used by NCQA, the data collecting organization requires participating providers to contract with audit vendors certified or licensed by the organization, who follow a standard auditing protocol.[41,42] In the **centralized approach**, used by the Centers for Medicare & Medicaid Services (CMS), the data collecting organization contracts with an independent entity to review a random sample of records across all providers. For Medicare, these audits are performed on a quarterly basis by a Clinical Data Abstraction Center (CDAC) and submitted to a data warehouse; a hospital's data are considered as "validated" if overall agreement with the reabstraction is at least 80% (http://qualitynet.org/).[43]

An auditing program was considered very important to the Better Quality Information (BQI) project (refer to Question 13), which sought to aggregate Medicare, Medicaid, and commercial claims data to assess physician performance. The BQI pilot sites created training and auditing processes to ensure accuracy, and the final measure results were given as feedback to providers who could challenge apparent errors or inconsistencies.[44]

### *Key Features for an Optimal Local/Regional Data Auditing Program*[41,42,45]

1. All **key data components of the measure should be audited**, including not only whether the numerator treatment or event occurred, but also whether the patient actually qualified for the denominator. Missing data are particularly important, because patients with missing data elements are typically excluded from quality reporting. Missing data rates should be tracked over time and across providers, so that high missing rates can be identified and corrected.
2. If the measure is based on an **adverse event** that is potentially susceptible to underreporting, such as a complication of care, then there should be **some effort through medical record review or linkage with other data (e.g., laboratory data) to find potential "false negative" cases, or unreported adverse events**. For the sake of efficiency, these efforts often focus on particular subsets of patients, such as those who were at very high risk of the adverse event.
3. Similarly, if a measure is based on a **specific process of care**, then there should be an effort to **find both "false positive" cases that were reported as having the treatment but did not, and "false negative" cases that were reported as not having the treatment but actually did**. For the sake of efficiency, these efforts often focus on particular subsets of patients, such as those who were reported as "exceptions" or as having unspecified contraindications to the standard treatment.
4. If possible, **auditing should occur concurrently with data collection** to detect errors in time to correct them before the data are used to support quality of care analyses. Record-specific feedback should be provided to submitting organizations to facilitate their review of reporting errors and their appeal of legitimate disagreements.
5. The auditing program should **verify that measure calculation processes conform to technical specifications**. This is most commonly done by creating a simulated data set or by manipulating an actual data set in a predetermined manner where the outcome is already known.

6. The auditing program should **assess system capabilities,** such as the ability to process information submitted by different provider organizations in different formats for consistent reporting of clinical measures.

## Question 6. How do HIPAA and other privacy regulations affect data collection and public reporting?

Community quality collaboratives may encounter opposition to the release of protected health information, from advocates of both patient privacy and physician privacy. In this environment, meeting State and Federal security standards for data sharing is critical to ensuring continued access to valuable data. This answer addresses the Federal issues around reporting, but community quality collaboratives are encouraged to learn about the specific privacy laws applicable in their State.

Adhering fully to all of the various privacy laws and regulations affecting data collection and public reporting can be quite complex, especially given the variability in State laws and the intricacies of the Health Insurance Portability and Accountability Act of 1996 (HIPAA). These laws and regulations dictate what and how health care information can be shared. Formal data sharing or business associate agreements must be in place prior to sharing or receiving protected health information. As described further below, those agreements must specify and limit how the data are used.

### *HIPAA*

Under the Administrative Simplification provisions of HIPAA, the Department of Health and Human Services (HHS) established national standards for electronic health care transactions and national identifiers for providers, health plans, and employers. HIPAA also addresses the security and privacy of health data.[46] To ensure privacy, HHS developed a set of regulations, commonly referred to as the HIPAA Privacy Rule, to address the use and disclosure of individuals' health information (called "protected health information") by organizations subject to the Privacy Rule, which are called "covered entities."[47] HIPAA defines "covered entities" as health plans, providers, and clearinghouses.

HIPAA further requires that covered entities have formal agreements in place with their business associates (e.g., a third-party pharmacy benefit management organization), which restrict the business associate to certain uses and disclosures. In addition, the American Recovery and Reinvestment Act of 2009, (sec. 13401) extends several privacy, security, and administrative requirements to business associates. Business associates will soon be required to comply with the same HIPAA requirements that apply to covered entities, and business associate agreements will need to be updated to reflect this requirement by early 2010.

### *Privacy Act of 1974*

The Federal Privacy Act of 1974 (Public Law 93-579) codifies the permissible personal information the Federal Government may collect and how it uses or discloses that information. The Privacy Act differs from HIPAA in several respects: it covers overall personal data collection and data use by the Federal Government only. HIPAA specifically targets the health care industry and restricts the sharing of protected health information with others, including the government.[48,49]

### *Resources for Creating Compliant Data Use Agreements*

Creating data use or business associate agreements between entities is necessary to meet the strict data privacy standards imposed by HIPAA. Several **Chartered Value Exchanges (CVEs)**, such as Wisconsin, Indiana, Washington-Puget Sound Health Alliance, and Minnesota, indicate that their business associate agreements meet all HIPAA and other privacy law requirements. These agreements may serve as prototypes that any community quality collaborative can follow; however, local legal advice may still be advisable to ensure full compliance with State privacy laws, which may be even more stringent than HIPAA. State **Quality Improvement Organizations** (QIOs), which contract with CMS to improve the quality and safety of health care for Medicare beneficiaries, are also offering assistance in creating data use agreements and business associate agreements that comply with HIPAA standards.[50] The **Centers for Medicare & Medicaid Services (CMS)** itself recently released guidance, in December 2008, as part of a toolkit "designed to establish privacy and security principles for health care stakeholders engaged in the electronic exchange of health information [that] includes tangible tools to facilitate implementation of these principles"[51] (http://healthit.hhs.gov/portal/server.pt). Finally, **statewide health data organizations** offer expertise in State health care privacy regulations, which can vary considerably by State.

### *Legal Considerations for Physician Tiering*

In recent years, several legal cases have addressed the ability of private entities to use enrollee-level claims data for tiering health care providers or public reporting on provider quality. In August 2007, Consumers' Checkbook/Center for the Study of Services won a Freedom of Information Act lawsuit in the U.S. District Court for the District of Columbia, which ordered CMS to release certain data from physician claims paid by Medicare, for the purpose of reporting the number of various types of major procedures performed by each physician and reimbursed by Medicare.[52] However, this decision was overruled by the U.S. Court of Appeals, which decided that physicians had a substantial privacy interest in not having claims data publicly disclosed because the data could be used, along with a publicly available Medicare fee schedule, to calculate a physician's total income from Medicare.[52]

This decision appears to establish, pending further court action, that physician-identified Medicare claims data are exempt from disclosure under the Freedom of Information Act. However, to the extent that community quality collaboratives engage in or facilitate efforts to create "tiered networks" of health care providers using quality and efficiency measures, they may face legal allegations regarding:

1. "Secrecy in both the standards used and the weights used to perform rankings;
2. The absence of a transparent rational basis for the methods chosen; and
3. The absence of a process by which physicians can examine the data on which their rankings rest and challenge errors in data or methodology."[52]

A recent legal settlement entered into by Regence Blue Shield and the Washington State Medical Association (WSMA), after a legal challenge by WSMA, exemplifies the importance of:

- **Inviting physician input into the data audit and methods used** to compare their performance;
- **Offering advance notice that new scores are forthcoming;**
- **Posting scores in an electronic format, along with an explanation of the methodology and data;**
- **Providing physicians the opportunity to appeal their scores**; and
- **Creating an independent external review process** to adjudicate these appeals.

This tension between protecting physician privacy, embodied in the Consumers' Checkbook decision, and enhancing fairness through transparency, embodied in the WSMA settlement, may play out differently in different markets.

Community quality collaboratives should note that they may encounter opposition to the release of health information from those concerned with physician and facility identification. Meeting State and Federal security standards for data sharing is critical, but in this environment, sensitivity to provider concerns is equally important for ensuring continued access to valuable data (refer to Question 24 about provider data reviews).

# Part II. Introduction to Measures of Quality

Data availability and validity are key elements to consider when selecting appropriate quality and resource use measures. It is also important to understand how measures are designed and constructed, how risk adjustment is performed, and how measure developers and endorsers are involved. Questions 7-13 introduce readers to a wide range of topics affecting measures of quality, as well as the national initiatives that are promoting the standardization of measurement.

## Question 7. How are quality performance measures constructed, and what are the implications of how their numerators and denominators are specified?

Quality performance measures are constructed in a variety of ways, including **proportions or percentages, ratios, means, medians, and counts**. Each approach serves a purpose and is appropriate in specific circumstances. Whichever approach is used, the detailed specifications and inclusion and exclusion criteria are typically developed through a painstaking process of discussion with clinical experts and analyses of empirical data. Measures with the same title, but sponsored by different organizations, may have somewhat different properties, as was recently demonstrated for a measure of hospital outcomes known as "failure to rescue."[52] Some indicators of hospital mortality exclude patients transferred in from other hospitals, whereas others include such patients.[53]

Minor but potentially confusing differences in the definitions of process-of-care measures should be reconciled, as The Joint Commission and the Centers for Medicare & Medicaid Services (CMS) have done for their Core Measures of hospital quality. Even the same measurement software, such as the AHRQ Patient Safety Indicators, can generate markedly different results depending on what may seem to be a minor choice, such as whether to turn on or off the option for using "present on admission" flags to identify events.[54] Accordingly, community quality collaboratives should be cautious in comparing results over time and across settings as measure specifications change.

### *Proportions and Percentages*

Most quality measures are constructed as proportions or percentages, where the denominator represents the number of persons treated by a health care provider during a defined time period who were at risk of, or eligible for, the numerator event. The **numerator then represents the number of persons in the denominator who received the appropriate diagnostic test or treatment** (e.g., aspirin for heart attack), or **the number who experienced an adverse outcome** (e.g., respiratory failure after surgery).

This method of constructing quality measures has several advantages, such as the fact that the range of performance is bounded between 0% and 100%, and the fact that multiple measures can easily be averaged to generate composite measures, as described in Question 10. This proportion/percentage method also facilitates comparison of performance across measures and sites. Its simplicity makes it understandable to consumers and actionable for health care providers; for example, most CAHPS® (Consumer Assessment of Healthcare Providers and Systems) survey questions on patients' experiences with care are transformed from their original form ("how often did your personal doctor….") to a dichotomous form ("always"/"usually" versus any other response), which can then be expressed as the percentage of patients with optimal or near-optimal experience.[55]

If multiple measures are presented side by side, then the polarity of some measures may need to be adjusted so that a higher percentage is always better (e.g., converting the percentage of patients who report a problem to the percentage who do not report that problem). **The major drawback of the proportion/percentage approach is that it ignores interesting variation among those who are categorized as "yes" or "no,"** such as the relative severity of a complication (e.g., bloodstream infections with or without sepsis), the relative importance of a patient's negative experience, or the timeliness with which an appropriate therapy was provided.

### *Ratios*

A few quality measures are constructed as ratio measures, in which **numerator cases may or may not be contained within the denominator**. For these ratio measures, the denominator is viewed as the best available proxy for the true population at risk, because that population cannot be enumerated. For example, the AHRQ Prevention Quality Indicators (PQIs) are expressed as hospitalizations per 10,000 residents of the target area per year, but the number of residents of the target area is estimated from previous Census data, which do not fully account for recent in-migration and out-migration.[56] The polarity of these measures must be clearly explained to consumers, because it is not immediately apparent whether lower values or higher values are better. In fact, cognitive testing has shown that consumers sometimes interpret higher asthma hospitalization rates as a sign of better care because they are concerned that aggressive health plans keep sick patients out of the hospital to save money.[57]

**The major drawback of this ratio approach is that the denominator may be a poor proxy for the true population at risk**. For example, only patients with diabetes are at risk for diabetes-related potentially preventable admissions, but the number of residents in the target area is a poor proxy for the number of diabetic patients. Given that the prevalence of diabetes varies across communities, PQI rates may vary for reasons unrelated to quality of care.

### *Means and Medians*

A few quality measures are constructed as **mean or median values**. For example, one widely used measure of emergency department/hospital care for patients with heart attack is the median time from arrival to administration of fibrinolytic therapy in eligible patients with ST-segment elevation or left bundle branch block on the electrocardiogram performed closest to arrival in an emergency department. The directionality of these measures must be clearly explained to consumers, because it is not always apparent whether lower or higher values represent better care. The advantage of this approach is that mean or median values capture subtleties of care, such as the timeliness of treatment, better than proportion or percentage measures. It may be possible to distinguish differences in performance using mean or median values that could not be distinguished using proportion or percentage measures. However, the drawback of this approach is that it makes the data more difficult to analyze and present, and it is not applicable to most quality measures.

### *Counts of Adverse Outcomes*

Finally, a few quality indicators are reported simply as **counts (i.e., number) of adverse outcomes, without any specification of the population at risk**. For example, the AHRQ Patient Safety Indicators for “Foreign Body Left In” and “Transfusion Reaction” are tabulated as counts at the hospital or area level, because they are extremely rare and every reported event merits investigation. These indicators are intended for surveillance purposes and not to compare performance across providers. Indicators of this type should not be used in public reporting or pay-for-performance programs, except for the limited purpose of promoting transparency.

## Question 8. What specific measures can be used to calculate physician performance at the individual or organization level?

Some physician leaders and physician organizations have long been vocal advocates of measuring physician performance. In the United States, Dr. Ernest Codman at Massachusetts General Hospital was a pioneer of this movement in the first two decades of the 20th century. Organizations such as the Society of Thoracic Surgeons and the American College of Surgeons have long maintained clinical registries to which physicians contribute data on their patient outcomes, and from which they can withdraw reports comparing their outcomes with external benchmarks. In 1991, the Department of Veterans Affairs launched its pioneering National VA Surgical Risk Study (NVASRS) in 44 VA medical centers, which evolved into the comprehensive program now known as the National Surgical Quality Improvement Project (NSQIP).[58] However, all of these efforts focused on confidential sharing of data through peer review mechanisms, which has been shown to improve patient outcomes but does not inform the market.[59]

### *Hurdles to Physician Performance Measurement*

Measuring physician performance for public reporting has been slow to take off due to **concerns about variation in patient risk at the physician level,[60] poor measure reliability, and limited or incomplete data for risk adjustment at the physician level**. For example, Scholle, et al., report that the denominator of eligible patients for a single physician from a single data source (i.e., health plan) is generally so small that results are unreliable.[61] In previous research, this problem applied even to one of the most common diseases in primary care, diabetes.[62] Awareness of these reliability issues is particularly important for pay-for-performance and public reporting programs, which are based on the hypothesis that performance varies meaningfully across physicians.[62]

### *Improving Completeness and Reliability*

Several options exist to address the small denominator dilemma: using composite measures; using group-level reporting; combining multiple years of data; and combining data sources (e.g., Medicare data and commercial carrier data). Some organizations are experimenting with **reporting composite measures** to enhance reliability, including HealthPlus in Michigan, which produces a public report of CAHPS® (Consumer Assessment of Healthcare Providers and Systems) and clinical measure composites by physician name (www.healthplus.org/PhysicianPerformanceReports.aspx). Kaplan and colleagues recently studied a national sample of 210 physicians with 7,574 diabetic patients participating in the NCQA-American Diabetes Association’s Diabetes Provider Recognition Program. They reported that process and intermediate outcome measures with a substantial "physician thumbprint" could be aggregated into a composite quality score with high reliability and excellent discrimination of physicians based on the quality of their diabetes-related care.[63] Additional research efforts in this area are now underway and will likely bear fruit in the next few years.

**Physician group-level reporting** is currently the most common approach to this problem. Several collaboratives (e.g., California's Integrated Healthcare Association, Wisconsin Healthcare Value Exchange, Massachusetts Chartered Value Exchange, and Washington-Puget Sound Health Alliance) publicly report the results of approximately 12 Healthcare Effectiveness Data and Information Set (HEDIS)-based measures related to diabetes, heart disease, asthma, preventive care, pediatric care, and depression, along with selected other measures. These measures were chosen because of the frequency of the underlying condition, the availability of national benchmarks, and the potential of the data to be available through the Centers for Medicare & Medicaid (CMS) Generating Medicare Physician Quality Performance Measurement Results (GEM) project. This is also the approach that has been adopted for Clinician and Group (C/G) CAHPS® reporting. However, physician group-level reporting suffers from numerous implementation problems, including the difficulty of assigning physicians who belong to multiple groups or who change groups, mandatory exclusion of physicians in solo or very small group practices, fluid group structures that may differ for different payers in the same market, and poor identification of consumers with physician groups (especially groups that have multiple sites).

Although **combining multiple years of data** may be an attractive option to improve the reliability of physician-level reporting, there is a serious tradeoff involved. As one reaches farther back in time to obtain sufficient data, one also loses the ability to make inferences about current or future performance. Most users find it untenable to reach back more than 3 years for historic data on quality. Such an ascertainment period may be sufficient for hospital-level measures,[64,65] and has recently been adopted by CMS for reporting on 30-day outcomes after hospitalization for myocardial infarction, heart failure, or pneumonia, but even 3 years of data are often insufficient for physician-level measures.

The final alternative to solving the small denominator dilemma involves **combining health care claims data from multiple payers**. This is perhaps the most attractive method for boosting reliability, because it eliminates the possibility of confusing consumers with conflicting information about the same care provided by the same physician during the same time period. As regional coordinators, chartered value exchanges (CVEs) and other collaboratives have the opportunity to drive this data collection process forward. At the national level, the Better Quality Information (BQI) project (discussed in Question 13) addressed the challenges with aggregating Medicare physician claims data and private payer data. This effort is now being moved forward by the Quality Alliance Steering Committee's Measure Implementation Work Group, which has selected Colorado and Florida as pilot sites for implementing "a nationally consistent data aggregation methodology" through a hub established by America's Health Insurance Plans Foundation. At the local level, the Wisconsin Collaborative for Healthcare Quality (which also served as one of the BQI pilot sites) has been a pioneer in collecting quality-related data from physician organizations.

### *Community Collaborative Example*

The **Wisconsin Collaborative for Healthcare Quality** (WCHQ) provides a unique example of physician organization measurement. Initiated 5 years ago by physicians who were motivated to improve the measurement system used by health plans, this collaborative includes most large medical groups and is starting to engage midsized groups as well. About 40% of physicians in Wisconsin submit clinical data from electronic records and chart review on 15 "home-grown" measures (derived from HEDIS measures) related to diabetes, hypertension, cardiovascular disease, postpartum care, and preventive screening. Their model embraces an all-patient, all payer philosophy. The denominator is derived through a three-question algorithm: "According to measure specifications, does the patient have the condition? Is this a patient who is managed by the group? Is this patient currently in the system?"

WCHQ also is piloting a **registry-based submission system (RBS)** for both WCHQ and CMS Physician Quality Reporting Initiative (PQRI) measures to make the data collection and validation process more efficient by allowing a few global patient files per reporting period to be submitted for aggregation. This also will expedite data validation. A revised business associate agreement permits secured patient-level data exchange for the transactions conducted by WCHQ and its associates.

### *Sources of Measures*

Combining the criteria outlined in Question 22 with the supporting information here and suggested measure sources (Table 4) can provide a basis for selecting specific measures of physician performance. The purpose of Table 4 is to provide an overview of **how the currently available measures are distributed across Institute of Medicine (IOM) domains and major source (also known as "developer" or "sponsor") organizations**. However, the number of available measures changes weekly, and the same measure may qualify for two or more domains. For example, almost any measure of effectiveness can serve as a measure of equity if it is used to compare performance across populations. There is overlap between measures of patient centeredness and timeliness, in that patients expect and deserve timely care. Therefore, the numbers in this table are presented to provide an overall view of current opportunities and challenges in physician performance measurement. Community quality collaboratives should take note of each measure's specification (e.g., age range, time period), which can differ between seemingly similar measures and substantially affect results.

The answer to Question 22 lists **several repositories that can be searched to identify physician performance measures in specific clinical domains**, across multiple developer or sponsor organizations. The most widely used of these repositories are the AHRQ's National Quality Measures Clearinghouse (http://www.qualitymeasures.ahrq.gov) and the National Quality Forum's list of endorsed standards (http://www.qualityforum.org/Measures_List.aspx). In addition, the AQA Alliance (described under Questions 19 and 21) offers a searchable compendium of approved performance measures (http://www.aqaalliance.org/performancewg.htm) that were submitted by at least five separate organizations, including some of those shown in Table 4.

**Table 4. National sources for specific physician performance measures**

| Physician Performance Measures | | | |
|---|---|---|---|
| **Measure Source and Description** | **IOM Domain**[a] | **Number of Measures Available**[b] | **Measure Example** |
| **1. AHRQ C/G CAHPS®**<br>Clinician and Group CAHPS® patient experience survey. Includes core questions for adult primary care physician (PCP), adult specialist, and child PCP surveys. Supplemental Items are also available. NQF endorsed.<br>Available at: www.cahps.ahrq.gov/content/products/CG/PROD_CG_CG40Products.asp?p=1021&s=213 | Safety | NA | NA |
| | Effectiveness | NA | NA |
| | • Process<br>• Outcome | NA | NA |
| | Patient Centeredness | A few primary composites in core survey and supplemental items that include numerous individual measures | *Composite*: How Well Doctors Communicate<br>Some *Individual Measures* included in the composite: Did the doctor listen carefully to you; show respect for what you had to say; spend enough time with you…? |
| | Timeliness | A few primary composite measures in core survey and supplemental items that include numerous individual measures | *Composite*: Getting Appointments and Health Care When Needed<br>Some *Individual Measures* included in the composite: Did you obtain an appointment for routine care as soon as you needed; receive an answer to your question the same day from the doctor's office; see the doctor within 15 minutes of your appointment time? |
| | Efficiency | NA | NA |
| | Equity | NA | The survey does capture sociodemographic information, which may permit some equity analysis of the other measures. |

| Physician Performance Measures | | | |
|---|---|---|---|
| **Measure Source and Description** | **IOM Domain[a]** | **Number of Measures Available[b]** | **Measure Example** |
| **2. CMS-PQRI[c]**<br>CMS established a physician quality reporting system, including an incentive payment for eligible professionals who satisfactorily report data on quality measures for covered services for Medicare beneficiaries. CMS uses 179 PQRI quality measures (2010) constructed by other measure developers (primarily NCQA and Physician Consortium for Performance Improvement® [PCPI], but also scientific groups such as the American Heart Association and American Stroke Association). Most are NQF endorsed or pending review.<br>Available at: www.cms.hhs.gov/PQRI/15_MeasuresCodes.asp#TopOfPage | Safety | About 20 | Medication reconciliation (within 60 days postdischarge) between discharge medications and current medication list in medical record from ongoing care physician<br>(Most safety measures are related to surgical care) |
| | Effectiveness<br>• Process<br>• Outcome | More than 100 | Asthma assessment for number of daytime or nocturnal asthma symptoms |
| | Patient Centeredness | Several | Consideration of rehabilitation services documented for stroke patients. |
| | Timeliness | More than 20 | Antidepressant medication prescribed during acute phase (first 12 weeks) for patients with new episode of major depression |
| | Efficiency | Several | Measures of overuse of potentially ineffective interventions (see NCQA and PCPI) |
| | Equity | NA | NA |
| **3. National Committee for Quality Assurance (NCQA)**<br>NCQA sponsors the *Healthcare Effectiveness Data and Information Set (HEDIS),* which includes measures of both health plan and physician performance. NCQA also sponsors Physician Recognition Programs for back pain, heart/stroke, diabetes, physician practice connections, and patient-centered medical homes. Most measures are NQF-endorsed.<br>Available at: www.ncqa.org | Safety | Several | Potentially harmful drug-disease interactions in the elderly, medication reconciliation postdischarge |
| | Effectiveness<br>• Process<br>• Outcome | More than 30 | Childhood immunization status, Immunization for adolescents, Breast cancer screening |
| | | NA | NA |
| | Patient Centeredness | Several as part of Back Pain Program | An ongoing system for obtaining feedback about patient experience with care, Use of shared decisionmaking |
| | Timeliness | Several | Followup after hospitalization for mental illness |
| | Efficiency | Several | Measures of overuse of potentially ineffective interventions, such as Avoidance of antibiotic treatment in adults with acute bronchitis |

## Physician Performance Measures

| Measure Source and Description | IOM Domain[a] | Number of Measures Available[b] | Measure Example |
|---|---|---|---|
| **4. PCPI**<br>The American Medical Association (AMA)-convened PCPI is involved in the development, testing, and maintenance of evidence-based clinical performance measures for physicians. At the time of publication, 111 measures were National Quality Forum (NQF)-endorsed and 150 were AQA Alliance approved.[66]<br>Available at: www.ama-assn.org/ama1/pub/upload/mm/370/measures.pdf | Safety | Several | Risk assessment for falls |
| | Effectiveness<br>● Process<br>● Outcome | More than 200 | Medication reconciliation (geriatric measure);<br>Stenosis measurement in carotid imaging reports (radiology measure) |
| | Patient Centeredness | Several | Advanced care plan (geriatrics) |
| | Timeliness | Several | Aspirin at arrival for acute myocardial infarction |
| | Efficiency | Several | Measures of overuse of potentially ineffective interventions, such as Avoidance of inappropriate use of antihistamines or decongestants, systemic antimicrobials, etc., for otitis media with effusion |
| | Equity | NA | NA |

[a] A single measure may belong to more than one domain. IOM definitions:
**Safety** – avoiding injuries to patients from the care that is intended to help them
**Effectiveness** – providing services based on scientific knowledge to all who could benefit and refraining from providing services to those not likely to benefit
**Patient centeredness** – providing care that is respectful and responsive to individual patient preferences, needs, and values
**Timeliness** – reducing waits and sometimes harmful delays for both those who receive and those who give care
**Efficiency** – avoiding waste, including waste of equipment, supplies, ideas, and energy
**Equity** – providing care that does not vary in quality because of personal characteristics such as gender, ethnicity, geographic location, and socioeconomic status
There are very few measures specifically designed to meet the IOM Equity domain criteria; however, Equity can be measured using almost any quality indicator if sociodemographic data are collected and linked to the indicator.

[b] Classification of measures is somewhat subjective in nature, so this column only approximates the distribution of measures across IOM domains. For example, the same measure of appropriate utilization could be interpreted as a measure of effectiveness or as a measure of efficiency.

[c] Most CMS PQRI measures were developed by other organizations, such as the AMA's Physician Consortium for Performance Improvement and the NCQA, so the same quality indicator may be referenced in multiple rows of this table.

## Question 9. What specific measures can be used to calculate hospital performance regionally or nationally?

Hospital performance measurement for public reporting has a longer history than physician performance measurement, with more established methods. Several organizations have been involved in developing and refining hospital performance measures over the last decade; including AHRQ (HCAHPS® [Hospital Consumer Assessment of Healthcare Providers and Systems] and Quality Indicators), the Centers for Medicare & Medicaid Services (CMS) (Quality Measures Management Information System), The Joint Commission, and the Leapfrog Group.

### *Sources of Measures*

Combining the criteria outlined in Question 22 with the supporting information here and suggested measure sources (Table 5) can provide a basis for selecting specific measures of hospital performance. The purpose of Table 5 is to **provide an overview of how the currently available measures are distributed across Institute of Medicine (IOM) domains and major source (also known as "developer" or "sponsor") organizations**. However, the number of available measures changes weekly, and the same measure may qualify for two or more domains. For example, almost any measure of effectiveness can serve as a measure of equity if it is used to compare performance across populations. There is overlap between measures of patient centeredness and timeliness, in that patients expect and deserve timely care. Therefore, the numbers in this table are presented to provide an overall view of current opportunities and challenges in hospital performance measurement. Community quality collaboratives should take note of each measure's specification (e.g., age range, time period), which can differ between seemingly similar measures and substantially affect results.

The answer to Question 22 lists **several repositories that can be searched to identify hospital performance measures in specific clinical domains**, across multiple developer or sponsor organizations. The most widely used of these repositories are AHRQ's National Quality Measures Clearinghouse (www.qualitymeasures.ahrq.gov) and the National Quality Forum's list of endorsed standards (www.qualityforum.org/Measures_List.aspx). In addition, the Hospital Quality Alliance (described under Questions 19 and 21) notes adopted measures (www.hospitalqualityalliance.org/hospitalqualityalliance/qualitymeasures/qualitymeasures.html) that were submitted by different organizations, including some of those shown in Table 5.

**Table 5. National sources for specific hospital performance measures**

| Hospital Performance Measures | | | |
|---|---|---|---|
| **Measure Source and Description** | **IOM Domain[a]** | **Number of Measures Available[b]** | **Measure Example** |
| **1. AHRQ HCAHPS®**<br>One of several surveys available through AHRQ. HCAHPS® surveys patients' experiences with their hospital care and services. NQF endorsed.<br>Survey tools and documentation resources are available at: www.cahps.ahrq.gov (Survey results are posted on CMS Hospital Compare Web site.) | Safety | NA | NA |
| | Effectiveness<br>● Process<br>● Outcome | NA | NA |
| | Patient Centeredness | A few primary composites that include numerous individual measures | *Composite*: Communication with Nurses<br>Some *Individual Measures* included in the composite: Did the nurses; treat you with courtesy and respect, listen carefully…? |
| | Timeliness | Several | How often did you get help as soon as you needed after pressing the call button? |
| | Efficiency | NA | NA |
| | Equity | NA | The survey does capture sociodemographic information, which may permit some equity analysis of other measures (although such data are not currently reported). |
| **2. AHRQ Quality Indicators**<br>Three sets of hospital QIs:<br>**A. Inpatient Quality Indicators (IQIs)** reflect quality of care inside hospitals, including inpatient mortality for medical conditions and surgical procedures.<br>**B. Patient Safety Indicators (PSIs)** also reflect quality of care inside hospitals, but focus on potentially avoidable complications and iatrogenic events.<br>**C. Pediatric Quality Indicators (PDIs)** reflect quality of care inside hospitals and identify potentially avoidable hospitalizations among children. | Safety | 25 Patient Safety Indicators<br>10 Pediatric Quality Indicators | ***PSI***: Hospital-level or area-level PSIs (e.g., foreign body left in during procedure, postoperative hemorrhage or hematoma)<br>***PDI***: Provider-level or area-level PDIs (e.g., foreign body left in during procedure, postoperative hemorrhage or hematoma) |
| | Effectiveness<br>● Process<br>● Outcome | 21 Inpatient Quality Indicators | ***IQI***: Mortality rates for medical conditions (7 Indicators) or surgical procedures (8 Indicators); Volume of procedures (6 Indicators) |
| | Patient Centeredness | NA | NA |
| | Timeliness | NA | NA |
| | Efficiency | NA | NA |

## Hospital Performance Measures

| Measure Source and Description | IOM Domain[a] | Number of Measures Available[b] | Measure Example |
|---|---|---|---|
| (A separate subset of **Neonatal Quality Indicators** is also available.)<br>Available at:<br>www.qualityindicators.ahrq.gov/ | Equity | NA | The data do capture sociodemographic information, which may permit some equity analysis of other measures. |
| **3. CMS-Hospital Compare**[c]<br>Publicly reports 24 clinical process-of-care indicators, 3 clinical outcome indicators, and 10 HCAHPS®-derived measures. Will begin reporting a subset of AHRQ QIs in 2010. National Quality Forum endorsed.<br>Available at:<br>www.hospitalcompare.hhs.gov | Safety | Several | Patients who received appropriate preventive antibiotics for their surgery |
| | Effectiveness<br>• Process<br>• Outcome | • More than 20<br>• Several | • Heart failure patients given an evaluation of left ventricular systolic (LVS) function<br>• 30-day mortality rate for heart failure |
| | Patient Centeredness | Several | 10 specific HCAHPS® questions |
| | Timeliness | Several | Pneumonia patients given initial antibiotics within 4 hours after arrival |
| | Efficiency | Several | 30-day readmission rate for heart failure |
| | Equity | NA | NA |
| | Equity | NA | NA |
| **4. The Joint Commission**<br>The Joint Commission accredits and certifies health care organizations and programs that meet certain performance standards. The Commission's core hospital measures overlap with many of the CMS hospital performance measures. NQF endorsed.<br>Available at:<br>www.qualitycheck.org | Safety | About 10 | Initial antibiotic received within 6 hours of hospital arrival |
| | Effectiveness<br>• Process<br>• Outcome | More than 30 | Angiotensin-converting enzyme (ACE) inhibitor or angiotensin receptor blocker (ARB) for left ventricular systolic dysfunction |
| | Patient Centeredness | Refer to HCAHPS | Refer to HCAHPS |
| | Timeliness | About 10 | Aspirin at arrival for heart attack patients |
| | Efficiency | NA | NA |
| | Equity | NA | NA |

### Hospital Performance Measures

| Measure Source and Description | Measure Source and Description | Measure Source and Description | Measure Source and Description |
|---|---|---|---|
| **5. Leapfrog**<br>The Leapfrog Group is a voluntary program aimed at rewarding the health industry for improvements in health care safety, quality, and customer value. Leapfrog designed quality and safety practice measures focused on reducing preventable medical errors. NQF endorsed.<br>Available at: www.leapfroggroup.org | Safety | Several | Patient safety practices include practices to prevent medication errors, to ensure appropriate intensive care unit staffing, and to refer patients to other higher volume hospitals for selected high-risk procedures. |
| | Effectiveness<br>● Process<br>● Outcome | Several | Specific processes of care for high-risk deliveries, weight loss surgery, aortic valve replacement, abdominal aortic aneurysm repair, pancreatic resection, esophageal resection |
| | Patient Centeredness | NA | NA |
| | Timeliness | NA | NA |
| | Efficiency | Several | Quality/cost ratings for heart bypass surgery, heart angioplasty, heart attack, and pneumonia care. Resource utilization is based on risk-adjusted average length of stay, inflated by 14-day all-cause readmissions. |
| | Equity | NA | NA |

---

[a] A single measure may belong to more than one domain. IOM definitions:
**Safety** – avoiding injuries to patients from the care that is intended to help them
**Effectiveness** – providing services based on scientific knowledge to all who could benefit and refraining from providing services to those not likely to benefit
**Patient centeredness** – providing care that is respectful and responsive to individual patient preferences, needs, and values
**Timeliness** – reducing waits and sometimes harmful delays for both those who receive and those who give care
**Efficiency** – avoiding waste, including waste of equipment, supplies, ideas, and energy
**Equity** – providing care that does not vary in quality because of personal characteristics such as gender, ethnicity, geographic location, and socioeconomic status
There are very few measures specifically designed to meet the IOM Equity domain criteria; however, Equity can be measured using almost any quality indicator if sociodemographic data are collected and linked to the indicator.

[b] Classification of measures is somewhat subjective in nature, so this column only approximates the distribution of measures across IOM domains. For example, the same measure of appropriate utilization could be interpreted as a measure of effectiveness or as a measure of efficiency.

[c] Most CMS HospitalCompare measures were developed by other organizations, such as The Joint Commission, so the same quality indicator may be referenced in multiple rows of this table.

## Question 10. What is the role and value of composite measures, and what are the most common approaches to constructing composites?

Composite measures, also known as summary or "roll-up" measures, combine individual measures into a single measure to summarize the overall quality of care delivered. AHRQ's *Talking Quality* Web site (www.talkingquality.gov/) defines **a composite measure as "condensing a number of quality measures into a single piece of information**."[67] This section presents the advantages and disadvantages of composites, background on composite construction, and considerations for scoring or weighting measures. Because different composite constructs and methods are appropriate for different purposes, the specific choices that community quality collaboratives make are less important than simply describing and providing some reasonable rationale for those choices.

### *Advantages and Disadvantages of Composite Measures*

Composite measures offer several important advantages over standalone measures, especially for public reporting and pay for performance:

1. They **reduce cognitive burden for consumers**, making it easier for sponsors to rank provider performance and for consumers to identify high-quality providers. They also minimize the danger of "cognitive shortcuts" that sometimes lead data users to make incorrect decisions when they are trying to interpret conflicting information from different measures.[68] For example, consumers may focus on one measure that they think is the most important, even if it is less informative than other measures.
2. They **enhance the reliability of quality measures**, which is especially important at the individual physician level, because it is typically very difficult to discriminate high-performing from low-performing physicians. This feature is also important for relatively rare outcomes, such as mortality from low-risk procedures.
3. They **fit well conceptually with pay-for-performance programs**, because the size of a provider's financial reward can be viewed as a composite measure of quality. Pay-for-performance rewards are typically based on multiple measures that are weighted and translated into dollar values. By creating and reporting composite measures, community collaboratives make this translation more explicit and set priorities to which providers can respond. However, when composites are used, providers often request to see their performance on each component of the composite so that they can decide how to concentrate their improvement efforts.[22]

Some potential disadvantages to using composite measures include:

1. **Difficulty achieving consensus on composite design and scoring**. Collaboratives may choose to use nationally vetted composites, such as those used in CAHPS® (Consumer Assessment of Healthcare Providers and Systems), to shortcut the laborious design process. Aside from these unusual examples, there is no professional consensus about how to construct and score composites.
2. **Loss of important information if the composite combines unrelated metrics**, thereby washing out meaningful differences on individual indicators (e.g., a hospital's performance on one or more specific indicators or procedures is significantly better or worse than its composite performance). Consumers may actually make the wrong

### *Producing Composite Measures*

Two different conceptual approaches or perspectives underlie most composite measures; each approach has its advocates and detractors.[69] The **psychometric perspective** is that an underlying, unmeasured factor, which we might call "quality," is the *cause* of what we observe with individual indicators. This approach is known as **reflective** because **the observed data reflect this underlying, unmeasured factor**, just as someone's IQ supposedly reflects his or her underlying intelligence. This approach requires a correlation *among* the measures included in the composite, because different measures can only reflect the same latent factor (i.e., quality) if they are correlated with each other. However, a problem with this approach is that different quality measures are often, in fact, uncorrelated or only weakly correlated with each other.[70,71]

The **clinimetric perspective** is not concerned about this lack of correlation; it uses clinical judgment rather than empirical analysis,[72] and it is intended to guide decisionmaking rather than to measure a mysterious, latent factor.[73] This approach is known as **formative** because the composite is ***formed from* or *defined by* specific indicators, through averaging.** For example, the Dow Jones Industrial Average is formed from market assessments of the value of 30 large corporations. This approach does not require any correlation among component measures. Although some authors still argue for testing composites to demonstrate their "internal consistency"[74] (avoiding what some have described as combining "apples and airplanes"[23]), others emphasize that "both approaches have a useful role to play."[69] On the next page, we discuss how a user's perspective affects his or her choice of a weighting method.

### *Steps for Constructing Composite Measures*

If a community quality collaborative chooses to construct composite measures, the following steps may be useful. These steps are further described in a recent report from the National Quality Forum on composite measure evaluation:

1. "**Identify the purpose** (e.g., comprehensive assessment of adult cardiac surgery quality of care) and delineate the quality construct to be measured (e.g., four domains of cardiac surgery quality include perioperative medical care, operative care, operative mortality, and postoperative morbidity).
2. **Select the individual measures and/or subcomposite measures** to be combined in the composite measure. This step may entail "standardizing" measures to have similar distributional properties so that they can be combined more easily.
3. **Ensure that the weighting and scoring of the components supports the goal** that is articulated for the measure. (Should the component scores be given equal weight or different weights based on some prioritization?)
4. **Combine the component scores**, using a specified method, into one composite (e.g., sum, average, weighted average, patient-level all-or-none scoring, etc.).
5. Finally, as with all measures, the composite requires **testing to determine if it is a reliable and valid indicator** of quality health care."

Using these criteria,[75] the National Quality Forum (NQF) endorsed three of AHRQ's four Quality Indicator composite measures: Mortality for Selected Conditions, Pediatric Patient Safety for Selected Indicators, and Patient Safety for Selected Indicators.

### *Reporting and Describing Composite Measures*

For community quality collaboratives developing their own composite measures, Kaplan and Normand provide specific recommendations about how to report and describe composite measures[23]:

1. "A clear and concise description of the **intended use of the composite** should be provided (including)… specific details regarding how the composite will be used to quantitatively measure provider performance."
2. "A **rationale** should be provided regarding the choice of the individual performance measures that comprise the composite (for example)… they are attributable to a provider; they vary across providers; they are mutable; and they are appropriate for measurement."
3. "A clear description of **how the items will be aggregated** is necessary… A method for handling missing data also needs to be articulated and justified."
4. "**Justification and definition of any case-mix variables** should be described."

### *Scoring Composite Measures*

The practical challenge in creating summary or "roll-up" measures is to decide how much weight to put on each component measure. **Under the reflective approach, composite developers typically use empirical (psychometric) methods** such as factor analysis and principal components analysis. Although such techniques are complex and require statistical expertise, they have the advantage of *automatically* generating weights that can be used to score composites. These weights usually reflect either the degree to which an individual measure explains an unmeasured or latent factor (e.g., quality) or its measurement reliability. Using the latter approach, more reliable measures with less random error are weighted more heavily because they are presumed to provide more valuable information.[76] In some cases, as for the Clinician and Group (C/G) CAHPS® survey,[77] sophisticated empirical methods have been used to validate relatively simple composites with equally weighted measures.[78,79]

**Community quality collaboratives that apply the formative approach must adopt their own weighting scheme** for scoring composite measures. Reeves, et al., reported that five commonly suggested methods for calculating composite measure scores, described in Table 6, resulted in very different scores for providers. They concluded that different methods are better suited to different types of applications.[80] For example, the "all-or-none" approach implicitly puts the most weight on the indicator with the poorest overall performance, so it is only appropriate when that indicator is actually the best "signal" of provider quality (or when overall performance is similar across indicators). The best candidates for all-or-none scoring are process measures "thought of as the indisputable basics of care for a given condition,"[81] because "the desired outcome depends upon completion of a full set of tasks" (i.e., partial execution is simply unacceptable).[80] **For most applications, these assumptions do not hold and equal opportunity or equal indicator weighting is more appropriate**.

The best approach, from the social perspective, might be to **weight individual measures based on their impact on population health**. For example, the AHRQ composite of “Patient Safety for Selected Indicators” is currently based on equal weighting of complications, although weights based on factor analysis (i.e., shared variance related to an underlying, unmeasured factor that might be called “quality”) have also been published. A future approach to this composite might assign weights based on the expected “return on investment” from preventing complications; for example, the marginal average impact of each type of complication on quality-adjusted life years,[82] hospital length of stay, or costs.[83,84] An alternative weighting scheme for the CMS process measures for heart attack and heart failure has recently been proposed, in which each measure’s weight is based on the product of its factor loading (to reflect its correlation with an underlying construct of quality) and its population standard deviation (to reflect its range for improvement). Compared with equal-opportunity weighting, this alternative scheme generates a composite that is more strongly associated with patient outcomes (i.e., inpatient survival).[85]

### *Community Collaborative Example*

The California Office of the Patient Advocate (OPA) used much of the above information to redesign how it publicly reports health plan performance data. For example, in addition to reporting the seven individual HEDIS diabetes measures, OPA created its own diabetes composite that rolled all seven indicators into a “topic score” called “Diabetes Care.” Similar composites were created for “Checking for Cancer,” “Chlamydia Screening,” “Treating Children,” “Maternity Care,” “Asthma Care,” “Mental Health,” “Heart Care,” and “Treating Adults: Right Care.” The design of these “topic scores” was informed both by empirical analyses of internal consistency and by structured input from consumers and other stakeholders. OPA selected “equal indicator weighting” and implemented an innovative approach to handle missing data, as explained in a technical document available on its Web site (http://opa.ca.gov/report_card/medicalgroupabout.aspx).

**Table 6. Types of scoring methods for composite measures (excluding empirically derived scoring methods)**

| Scoring Method for Composite | Definition | Example | Adopting Organization |
|---|---|---|---|
| **All or None** | The percentage of patients for whom all indicators triggered by that patient are met. | “Appropriate Care Measure” for 4 conditions (8 heart attack, 4 heart failure, 7 pneumonia, and 5 Surgical Care Improvement Project treatments [per Joint Commission]). Patients must receive *all* identified interventions for a condition to be included in the numerator. | Pennsylvania Health Care Quality Alliance’s Progress and Performance Report of Hospital Quality |
| **70% Standard** | Based on all or none with less strict criteria; 70% or more of indicators must be triggered. | None to our knowledge (but hypothetically attractive). | |
| **Overall Percentage (Opportunity Weighting)** | Percentage of all audited care events that are met, where each opportunity to “do the right thing” counts equally. | 149 hypertensive patients triggered 26 hypertension indicators 828 times. Required care was given 576 times, yielding 69.6% (576/828). | Centers for Medicare & Medicaid (CMS) Pay for Performance (P4P) Premier Hospital Quality Incentive Demonstration Project[86] |
| **Indicator Average (Equal Indicator or Event Weighting)** | For each indicator, the percentage of times the indicator was met is computed and scores are averaged across all indicators. This represents the mean rate at which each aspect of care was met. | Hospital quality of care for acute myocardial infarction, congestive heart failure, and pneumonia is rated by averaging multiple indicators within each clinical domain. | Hospital Quality Alliance (HQA) |
| **Patient Average (Equal Patient Weighting)** | The percentage of triggered indicators successfully met is computed for each patient. Scores are averaged across all patients to represent the mean percentage of audited aspects of care met for each patient. | None to our knowledge (but hypothetically attractive). | |
| **Expert Opinion (Evidence Based)** | Each indicator is weighted based on evidence of its impact on population health or evidence of the effort required to achieve it. | General Medical Services contract pays physicians more for achieving performance targets that require more time and other resources. | United Kingdom National Health Service |

**Source:** Definitions and examples in this table are summarized and adapted from Reeves, et al.,[80] and the Pennsylvania Health Care Quality Alliance.

## Question 11. What is "risk adjustment" and how is it best applied?

Risk adjustment involves using statistical methods to "level the playing field" by adjusting for the effects of patient characteristics that may vary across providers. Without risk adjustment, users can easily draw incorrect conclusions, because the hospitals or physician organizations that appear to have the worst outcomes may simply have the sickest patients. Risk adjustment is particularly important for outcome measures, because patient outcomes are driven not just by quality of care but also by age, gender, medical history, comorbid illnesses, behavioral and social factors, and physiologic factors. Risk adjustment is not used for structural measures of quality, such as whether hospitals have implemented appropriate error prevention practices, according to the Leapfrog Safe Practices Survey, because implementation of these desirable structures is not related to patient characteristics.

### *Limitations to Risk Adjustment*

The major limitation of risk adjustment is that **it can only account for measurable and reported risk factors**. Unfortunately, many important risk factors for adverse patient outcomes are either not measurable using available data (e.g., preoperative functional status) or are not consistently reported (e.g., obesity). For some outcome measures, such as heart attack mortality at the hospital level, classification of hospital performance is reasonably robust despite these immeasurable or unreported factors, because between-hospital variation in outcomes is relatively large.[87,88] In addition, unmeasured risk factors tend to be randomly distributed across hospitals.[89] For other outcome measures, the limitations of risk adjustment are likely to be more problematic, particularly at the physician level, due to clustering of certain types of patients in certain physicians' practices.

Another problem with risk-adjusted outcomes is that they are often misinterpreted. Most risk-adjustment approaches involve estimating **indirectly standardized outcome ratios**, also referred to as ratios of observed to expected outcomes. These ratios compare the actual outcomes of the specific set of patients treated at each hospital with their expected outcomes had they been treated by an average hospital in the population. If a hospital is identified as a poor outlier, then its outcomes were significantly worse than what would have been expected if the *same patients* had been treated at a hypothetical average hospital. In other words, each hospital is compared with the hypothetical average hospital *treating the same patients*, not with any specific hospital *treating different patients*.[90] Therefore, **community quality collaboratives should avoid ranking hospitals based on their risk-adjusted outcomes, even though such rankings are easy for users to interpret.** If Hospital A's outcomes are significantly better than expected, while Hospital B's are not, then we are more confident that Hospital A offers high quality of care, but we cannot assume that Hospital A is actually better than Hospital B. Therefore, hospitals should be placed into a limited number of "bins" (typically 3-5) based on statistical criteria, and ordered alphabetically or geographically (not ranked) within those categories.

### *Risk-Adjustment Options*

Risk adjustment may be implemented in a wide variety of ways, but most community quality collaboratives use one of the following approaches:

1. **Adopt indicators that have already been risk adjusted by an intermediary** (e.g., Centers for Medicare & Medicaid [CMS] measures of 30-day mortality after heart attack, heart failure, or pneumonia).[91]
2. **Use off-the-shelf methods that are built into readily available software programs** (e.g., AHRQ Inpatient Quality Indicators and Patient Safety Indicators, 3M Health Information Systems' all patient refined-diagnosis related groups (APR-DRGs), Chronic Disability & Illness Payment System for the Medicaid population).
3. Some community quality collaboratives use risk-adjusted rates calculated by State health agencies that **develop their own risk-adjustment models** for selected conditions or procedures, such as coronary artery bypass mortality in Massachusetts, New York, New Jersey, California, and Pennsylvania. This type of customized modeling is most important when community quality collaboratives want to take advantage of particular strengths of their local data, such as "present on admission" coding of every diagnosis in California (and other States) and "key clinical findings" in Pennsylvania. Both of these data features dramatically improve risk adjustment[92]; adding fewer than 15 laboratory findings has been shown to eliminate more than 75% of the estimated bias in hospitals' expected mortality rates for major medical conditions.[93]

Community quality collaborative that are interested in customized modeling, similar to what has been done in California and Pennsylvania, should refer to a standard text in the field before undertaking such analyses. These texts explain the standard methods for estimating, assessing, and validating customized models.[94]

**Community quality collaboratives may not have access to the data needed for risk adjustment, even when risk adjustment is desirable**. For example, HCAHPS® (Hospital Consumer Assessment of Healthcare Providers and Systems) survey results regarding hospital care are adjusted for the effects of both mode of survey administration and patient mix before they are publicly reported by the Hospital Quality Alliance. Generally speaking, HCAHPS® adjustments for survey mode are larger than adjustments for patient mix.[95] The factors included in patient-mix adjustment include respondent education, age, self-rated health status, emergency room admission, primary language, and service line (i.e., maternity, surgical, medical).[96] Although risk adjustment is also useful for patients' assessments of health plans and clinicians,[97,98] and the CAHPS Analysis Programs downloadable from the AHRQ Web site include optional risk adjustment (www.cahps.ahrq.gov/cahpskit/CAHPSKIT_main.asp), most community quality collaboratives do not receive the respondent-level data necessary for risk adjustment.

### *Alternatives to Risk Adjustment*

**Not all quality measures require risk adjustment**. Two alternative approaches, which are more commonly used for process measures, include risk stratification and exclusion. Under **risk stratification**, patients are divided into two or more groups according to their expected risk of the process or outcome of interest. For example, CMS's Nursing Home Compare system includes the "Percentage of High-Risk Long-Stay Residents Who Have Pressure Sores" and the "Percentage of Low-Risk Long-Stay Residents Who Have Pressure Sores" as separate measures (www.medicare.gov/NHCompare/). Occasionally, stratification is used to support numerator definitions of process measures that differ according to the patient's risk status. In this case, a broader time window for the process measure is allowed if the patient is classified as low risk.

Risk stratification can be applied to reporting of CAHPS® data when risk adjustment is impossible; for example, plans could be asked to report separately on the experiences of healthy and sick members,[99] members in different markets,[100] or members with different benefit designs. Stratification may be particularly helpful for exposing disparities in care and for rewarding plans and physician groups that reduce disparities.[101,102] However, **reporting stratified data typically requires larger sample sizes than reporting aggregated data**, or else stratum-specific estimates of performance are unreliable. Community quality collaboratives may need to provide additional resources to support collecting and reporting stratified data at the local level.

A more widely used approach, however, is simply to **exclude patients who do not qualify for the process of care in question**, or for whom the process of care has not been shown to confer a clear benefit. For example, all of The Joint Commission's Core Measures of hospital quality for heart attack, heart failure, pneumonia, and surgical care have carefully defined denominators that exclude patients for whom the therapy in question is documented as medically inadvisable (www.jointcommission.org/PerformanceMeasurement/PerformanceMeasurement/default.htm).

### Question 12. What are the opportunities and challenges to using patient experience surveys to measure hospital or physician performance at the regional or State level?

Both clinical treatments and patient experiences are important facets of the overall quality of care. In the absence of a standardized set of tools to assess patient experience, many providers in the 1990s designed their own surveys or contracted with leading vendors (e.g., Press Ganey, PRC) to administer vendor-specific surveys. Additionally, Web-based patient experience sites, such as Vitals.com and AngiesList.com, have been established. Some of these sites report licensure/certification information as well as the opinions of a nonrepresentative sample of patients who initiate a posting. Although these sites may demonstrate consumers' desire for data to inform their health care decisionmaking, there is generally no scientific rigor supporting the conclusions.

Today, the premier tool for measuring patient experiences with care is the Consumer Assessment of Healthcare Providers and Systems (CAHPS®) series of surveys created by AHRQ (www.cahps.ahrq.gov). This standardized survey series has been constructed carefully, tested rigorously, endorsed by the National Quality Forum, and accepted by stakeholders nationally. The two surveys most relevant to community quality collaboratives are the **Hospital CAHPS®** and the **Clinician and Group (C/G) CAHPS®**. In addition, the **National CAHPS® Benchmarking Database (NCBD)** provides national benchmarks for many of the surveys.

### Hospitals: HCAHPS

Hospitals contract with a Centers for Medicare & Medicaid (CMS)-approved vendor to administer the HCAHPS® survey to a sample of all inpatients (not just Medicare beneficiaries), which makes the results relevant to community quality collaboratives. To encourage participation, CMS now links annual hospital payment updates to submission of HCAHPS® survey results. The results are publicly reported on the CMS Web site: www.HospitalCompare.hhs.gov. Some CVEs, such as the Washington-Puget Sound Health Alliance and the Maine Chartered Value Exchange Alliance, incorporate the CMS HCAHPS® results into their public reports.

### Physicians: C/G CAHPS

In 2009, the NCBD will provide preliminary benchmarks for the Clinician-Group survey, which will offer useful comparisons for CVEs and local collaboratives already measuring patient experience with physician care. Preliminary evidence from CAHPS® surveys indicates that providing feedback to physicians and their practices improves their quality of care.[103] Supplemental items are now being developed to address issues of particular concern to the Medicaid and Children's Health Insurance Program (CHIP) populations, including care for children with chronic conditions and people with impaired mobility, reduced health literacy, and other special health care needs.

### Patient Experience Surveys: Challenges and Possible Solutions:

1. **Challenge: Cost of survey administration**
   **Possible Solution:** The resource challenges of administering the C/G CAHPS® survey for small group practices can generally be overcome. A study by a small practice in Pennsylvania found that for about $1,000, or $200 per physician, reliable data could be captured.[5] The CAHPS® Consortium estimates a cost of $8 per completed survey for mail administration, or $360 per clinician. Practices of all sizes can obtain data appropriate for benchmarking and goal setting from the NCBD. The CAHPS® User Network recommends that practices contact researchers at local universities for help with statistical analysis, although a better long-term strategy may be to create a permanent infrastructure for data management and analysis through CVEs or similar collaboratives.
2. **Challenge: Difficulty of gaining provider buy-in**
   **Possible Solution:** Some pay-for-performance programs reward providers for participating in the survey. For example, California's Integrated Healthcare Association (IHA) sponsors a pay-for-participation incentive program to encourage physicians to participate in California's version of C/G CAHPS®. Nearly universal hospital buy-in with HCAHPS has been achieved through a 2% annual payment update incentive from CMS.
3. **Challenge: Concerns about differences in case mix across providers**
   **Possible Solution:** Transparent methodologies for case-mix adjustment have been carefully tested by the CAHPS® Consortium. The CAHPS® Consortium recommends adjusting for self-reported general health status (i.e., excellent, very good, good, fair, poor), age, and education. Older individuals and those in better health tend to rate their care, plans, and providers higher than younger individuals and those in worse health. There is also evidence from a number of studies that education affects ratings, with more educated individuals giving lower ratings. However, users of CAHPS® Analysis software

can specify an unlimited number of adjuster variables or choose not to adjust the data at all, depending on their preferences and data quality.

4. **Challenge: Duplication of data collection effort**
   **Possible Solution:** Another barrier to physician participation is the duplication of effort that occurs when a single physician's patients are surveyed multiple times by different organizations (e.g., health plans, State agencies). To minimize that redundancy and use resources efficiently, AHRQ is developing a strategy for administering the survey for a physician only once in a given period.[53] California avoids redundancy through its California Cooperative Healthcare Reporting Initiative (CCHRI), which coordinates the annual survey of medical groups statewide. They also coordinate public reporting of results with the State's Office of the Patient Advocate. Massachusetts Health Quality Partners, part of the Massachusetts Chartered Value Exchange, coordinates a similar effort.[104] Others are considering administering the survey biennially.
5. **Challenge: Poor response rates**
   **Possible Solution:** In general, response rates to mail surveys have been declining over the past two decades. Variation in response rates across provider organizations may lead to bias in estimating comparative performance. Three modes of survey administration are available (mail only, telephone only, mixed mode); mail with telephone followup achieves the best response rates (i.e., about 40%, or about 10% higher than mail only) in most settings. Regardless of which mode is selected, survey sponsors should always follow recommended protocols to improve contact rates and response rates and should report their results. According to CAHPS® reports, patients may give more positive reports and ratings of care when the data are collected by telephone as opposed to mail, but there is not yet a standard method to adjust for this difference.

### *Community Collaborative Example*

In a nationwide effort to report patient experience with physician care, **Consumers' Checkbook/Center for the Study of Services** is testing a modified version of C/G CAHPS with three collaboratives in a pilot project to report patient experience at the individual physician level. One of the collaboratives, the **Kansas City Quality Improvement Consortium**, reported that this pilot survey is very similar to the C/G CAHPS® survey, except for a few "dropped" questions and a few modified demographic questions. Consumers' Checkbook worked with the local communities and health plans to identify the patient and physician population. Of the seven health plans in the area, three (including the largest) contributed patient and physician contact information.

A community awareness campaign about the survey was implemented to familiarize physicians and consumers with its purpose and to increase response rates. Surveys were mailed in November 2008 and again in January 2009 to nonrespondents. They achieved a 47% response rate in Kansas, with an average of 57 responses for each of the 713 participating physicians. Preliminary results will be distributed to physicians for a 60-day review and comment period, as required by NCQA. The goal was to create a useful presentation of information, based on consumer focus group feedback and to publicly report results in 2009.[105]

## Question 13. What is the "Better Quality Information" pilot project, sponsored by the Centers for Medicare & Medicaid Services, and what can be learned from it?

The **Better Quality Information to Improve Care for Medicare Beneficiaries (BQI) Project** (www.cms.hhs.gov/bqi/) is a CMS-funded pilot project that ended in late 2008. The Delmarva Foundation, CMS's contractor, subcontracted with six pilot sites, five of which were Chartered Value Exchanges (CVEs), to test methods to aggregate Medicare claims data with claims data from commercial health plans and Medicaid to calculate and report quality measures for physician groups and individual physicians. (The pilot sites were California Cooperative Healthcare Reporting Initiative, Indiana Health Information Exchange, Massachusetts Health Quality Partners, Minnesota Community Measurement, Phoenix Regional Healthcare Value Measurement Initiative, and Wisconsin Collaborative for Healthcare Quality.)

### *Project Aims*

The project aims were: (1) to study the challenges and benefits of aggregating Medicare fee-for-service data with other regional quality data, including commercial payer administrative data and provider-submitted data, to calculate quality measures for ambulatory care; (2) to study the benefits of reporting quality measures to physicians and other providers of care; and (3) to study the benefits of reporting quality measures to beneficiaries.

### *Findings Pertinent to Measure Selection*

The final report provides a rich source of information that can guide the selection of measures and ensuing data collection.[44] In particular, CVEs and other collaboratives may find Chapter 2 relevant to their needs in selecting measures. The BQI project applied a version of the measure selection criteria outlined in Question 20 of this Decision Guide, including an iterative process that considered several measure sets authored by different developers.

The six sites first considered the AQA Starter Set and later considered Healthcare Effectiveness Data and Information Set (HEDIS) and other measure sets. Their reasons for excluding individual measures ranged from limited relevance to Medicare beneficiaries (e.g., screening for HIV) to the need for medical chart review, which was outside the scope of this project. Three pilot sites chose to retain locally developed measure specifications for the selected BQI measures. The sites also varied in their use of data sources, in how they defined their target population, and in how they included medical professionals. The BQI report states that they were "encouraged by both the consistency and the variation" in measures as it allowed analysis of the effects of minor differences on measure outcomes.

### *Findings Pertinent to Data Sources and Attribution*

Most of the sites included administrative data from commercial payers, and many included clinical data reported from physicians, hospitals, and other providers. Most sources were electronic but some were from paper records. Some sites had experience with electronic transmission of laboratory data and pharmacy/Medicare Part D data. Generally, their choices were guided by what was available and practical.

Proper attribution of patients (to physician and group) and physicians (to medical group) is critical to ensuring that physician and group scores are calculated and interpreted correctly. Attribution proved to be challenging because physician identifiers (Unique Physician

Identification Numbers [UPINs] and Taxpayer Identification Numbers [TINs]) were either not consistently available on all claims or could only link patients to large, corporate groups, rather than clinic sites.

**All three methods (UPIN, TIN, and physician group roster) of assigning physicians into groups using claims data were found to have inaccuracies of 10% or more**. The project participants concluded that it is preferable to aggregate data at the individual provider level. Results can “then be combined using consistent rosters for medical group-level reporting.” However, merging data sets at the higher level of medical groups introduced complex errors due to nonstandard assignment of physicians to groups. The report stated, “assigning physicians to groups using TINs is easiest because the TIN is available in all encounter forms while group identifiers and UPINs are not. Methods to allow individual providers to correct their medical group memberships were found to be effective by the BQI pilot sites” (Appendix 6 of the BQI report provides detail).

### *Conclusion*

All six pilot sites thought that **inclusion of Medicare fee-for-service data gave “a more complete picture of care quality”** because Medicare beneficiaries represent a major population segment in all pilot communities. Moreover, increasing the “N” (denominator) for individual physicians helped to stabilize the measure results, which would not have been feasible without the Medicare data. **Harmonizing data standards and measure specifications** is critical to meeting regional and national comparative reporting needs, and the BQI project contributed to accomplishing that difficult task. **In future followup projects, proper attribution of claims to individual physicians or physician practice sites is the critical issue** that will need to be addressed.

# Part III. Introduction to Resource Use/Efficiency Measures

Resource use measurement is in the early stages of development. Although public and private payers express considerable interest in calculating the value of health care services, it remains a challenge to develop and implement nationally accepted measures. Questions 14-18 describe both theoretical (e.g., types of measures, measure construction) and applied (e.g., measure use in comparing providers, national efforts) aspects of resource use measurement.

## Question 14. What are the main types of resource use measures?

The term "resource use measures" is intended to broadly capture indicators of the cost and efficiency of health care provision. **Health care resource use measures reflect the amount or cost of resources used to create a specific product of the health care system**. The specific product could be a visit or procedure, all services related to a health condition, all services during a period of time, or a health outcome. **"Efficiency" measures are a subset of resource use measures that compare the production of products of a specified level of quality**.[1,106] Most resource use measures in use are not efficiency measures by this definition because they do not explicitly incorporate a measurement of the quality of the product.

A systematic review of available resource use measures was published by AHRQ at www.ahrq.gov/qual/efficiency/index.html. Three main groups of resource use measures have been developed:

1. **Relatively simple measures of the resources used to produce health care**, such as mean length of stay, mean charges or estimated costs, and readmission rates for hospitals; and consultation or test ordering rates for outpatients with common complaints such as low back pain.
2. **More complex measures of health care resource use**, including both inpatient and outpatient services, using econometric or mathematical programming techniques to account for multiple outputs.
3. **Measures of the resources used in an episode of care for a patient, or to treat a patient with a specified burden of comorbidity for a specified period of time**.

### *Relatively Simple Measures*

The first group of measures includes relatively straightforward measures long used in hospital management. The most common measure of this type is the average length of hospital stay, adjusted for case mix. This provides an estimate of the resources used to care for a hospitalized patient with a particular diagnosis. Other measures focus on whether the hospitalization itself was a necessary use of resources; potentially avoidable readmissions following hospital stays are a commonly used measure of this type. Finally, charges or estimated costs associated with specific services are sometimes presented as resource use measures, although these measures may be distorted by cost shifting, anticompetitive behavior, differences in quality, and a variety of other manifestations of market failure.

### *More Complex Measures*

The second group of measures, typically published in the peer-reviewed literature,[107] reflects the amount and type of various resources used to produce a mix of hospital services, such as hospital discharges, outpatient visits, and procedures. These measures use complex methods to account for different mixes of resources used and services produced. The complexity of these methods may have inhibited the broad use of these measures beyond academic research, because measurement results can be sensitive to many specification choices and difficult to interpret. However, a measure of this type is included in AHRQ's *National Healthcare Quality Report* (refer to Question 17). Using a related approach called Data Envelopment Analysis (DEA),[108] Valdmanis and colleagues compared the number of hospital staff and beds used to produce a mix of inpatient and outpatient services across 1,377 urban hospitals in 34 States operating in 2004. Their study found that hospitals could increase the total amount of outputs produced by an average of 26%, without increasing inputs, by eliminating inefficiency.

### *Episode- and Population-Based Measures*

The third group includes two main approaches to resource measurement: (1) **"episode-based"** measures of resources used for an "episode of care," including all services related to a particular medical condition or acute event; or (2) **"population-based"** measures of resources used in providing all care to an individual with one or more chronic conditions for a period of time. Of the two approaches, episode-based measures have been used most widely by commercial payers and have been recommended for use in Medicare by the Congressional Budget Office and the Medicare Payment Advisory Commission, among others.

Episodes are defined using "grouper" tools, such as the Episode Treatment Groups (ETGs) developed by Symmetry Health Data Systems and Medstat Episode Groups (MEGs) developed by Thomson Medstat. These tools group related services into episodes primarily using diagnosis codes; episodes include services furnished by different providers in different care settings. The cost or resources used to produce each episode are then tallied across providers.

A population-based approach to efficiency measurement, such as Diagnostic Cost Groups (DCGs), classifies a patient population according to morbidity burden in a given period (e.g., one year). The cost or resources used for all health care for that patient over the time period are then measured.

## Question 15. What types of data are used to construct resource use measures? How is "cost" measured?

**Resource use measures are typically constructed using administrative data**. Hospital-focused measures can use administrative data sets such as those collected and disseminated by statewide health data organizations.[109] Measures that cover a broader range of services and care settings may require the use of insurance claims for medical, ancillary, and pharmacy services. While administrative data do not include much clinical information, they have the advantage of being readily available and reasonably standardized. Some measures append additional data from other sources on provider characteristics, such as the American Hospital Association Annual Survey.

One of the main challenges in using administrative data is that each insurer's data include only a portion of all care provided by each provider. **Data from one insurer may not be sufficient to allow for stable measurement or may not be representative of a provider's entire practice**. For this reason, some initiatives have aggregated data across multiple sources. For example, statewide health data organizations collect all-payer hospital data, and six communities have aggregated data from multiple sources for the Centers for Medicare & Medicaid Services (CMS) Better Quality Information project, described under Question 13.[44] However, aggregation of cost data can be both technically and politically challenging, because insurers and providers are reluctant to share sensitive information, such as pricing arrangements.

### *Alternative Sources for Capturing Cost*

For a profit-maximizing firm in a perfectly competitive market with symmetric information, the marginal cost of producing a service equals the marginal revenue or transacted price of that service. However, **health care markets are not perfectly competitive, and transacted prices in the commercial market are generally unknown**. Therefore, there are two easy alternatives to using actual prices (payments) in measuring hospital costs. One option is to use the **amounts that providers charge payers**, which are more readily available than either prices (payments) or hospital costs. However, prices (payments) often differ significantly from charges due to negotiated discounts, bundling of claims, and shared risk or capitated payment. For hospitals, **charges can be used to estimate hospital resource costs by applying cost-to-charge ratios** calculated from Medicare hospital cost reports or similar all-payer systems established by several States (e.g., California, Florida, Massachusetts, New Jersey). Although the accuracy of these estimates at the service level has been questioned, they appear to have reasonable validity at the hospital level.[110,111]

Another alternative, most commonly applied to ambulatory care, is to **use standardized units to assign the relative resource use of different services**. For example, relative value units (RVUs), which are used by Medicare and other payers to determine relative payment rates for various procedures, could be used instead of the price paid for services. One group that has followed this approach is the Washington-Puget Sound Health Alliance. The Alliance has constructed a regional all-payer database that is being used for quality and resource use measurement, but the data suppliers do not submit any financial information. Instead, The Alliance uses a system of RVUs developed by Milliman to score different services (including not only physician services but also other types of services without Medicare RVUs) using a common metric for relative resource use.[112]

Different questions are answered by using standardized prices, which are the same for all providers and payers, than by using charges or actual prices, which differ across providers and payers. **Standardized prices address whether a health care service could be produced faster, with fewer people, fewer labor hours, or fewer supplies** (i.e., fewer inputs). **Charges or actual prices address whether the output could be produced less expensively** (i.e., reducing the total cost of labor, supplies, and capital, either by using fewer inputs or by procuring those inputs at lower cost).

Consumers and purchasers may favor using actual prices, which reflect what they actually pay. For example, the price of an imaging study is likely to be higher at a teaching hospital than at a community hospital, even though the same real resources may be used in each setting.[113] In this hypothetical example, a measure based on actual prices would reflect the fact that consumers and payers pay more for the imaging study at a teaching hospital than at a community hospital. By comparison, a measure based on standardized prices might show that the same quantity of resources was used for the imaging study in both settings.

## Question 16. What is known about the validity of available resource use measures, including their advantages and disadvantages?

The state of the art in health care resource use measurement contrasts sharply with that of the measurement of health care quality. **Little is known about the validity of resource use measures or the advantages and disadvantages of different measures**. Only a few resource use measures (length of stay and readmission measures) have been endorsed by the National Quality Forum (NQF). Unlike the evolution of most quality measures, current resource use measures are not typically derived from practice standards in the research literature, professional medical associations, or expert panels. Unlike most quality measures, resource use measures have been subjected to few rigorous evaluations of their reliability and validity.

### *Differences Among Resource Use Measures*

Several differences among resource use measures could guide a community collaborative's choice of measures. Many resource use measures focus on hospitals, including simple measures such as mean length of stay and more complex multiple-output measures using econometric or mathematical programming techniques. Hospitals account for a high proportion of total health spending, and so may be of particular interest for resource use measurement. However, a focus on hospital care omits many types of services and does not capture coordination of care across settings, where many inefficiencies in delivery occur.

Commercial measures, both episode based and population based, reflect care provided in multiple settings. Population-based measures, although typically adjusted for patients' risk of higher resource use, reflect the "probability risk" that patients will acquire a condition that requires higher than expected resources during the data collection period. Episode-based measures, in contrast, reflect only the resources used in treatment of a particular condition, beginning at the onset of the episode of care for that condition.

After weighing these considerations, **several national groups, including NQF and CMS, have expressed a preference for episode-based measures over population-based or hospital-focused measures**. CMS lists the following advantages of episode-based measures[114]:

- "Compare more similar patients than per capita calculations, as they are defined by similar procedures or conditions;
- Capture the multiple ways in which services can be combined and substituted to produce the best outcome at the lowest cost;
- Reflect patients' view of care as they move between and across settings and managers of their care, rather than simply measuring resources used for just a part of their care in one setting, and
- Encourage improved coordination across settings included in the episode."

Many hospital-focused measures, such as average length of stay and readmission rates, are widely used and relatively simple to construct. For example, United Health Group sponsors an NQF-endorsed measure described as "overall inpatient 30-day hospital readmission rate."

However, “single output” measures of this type may be misleading, because the services needed to avert readmissions (e.g., longer inpatient stays) may actually consume more resources than the “preventable service” itself. Readmission may be an undesirable outcome for some patients in some settings, but a desirable outcome for other patients in other settings.

More complex multiple-output hospital measures are published in the peer-reviewed literature. However, they are generally published in one-off studies and use complex methodology, so that community quality collaboratives would need to reconstruct such measures at considerable cost. Commercial episode-based and population-based measures are proprietary and are available to be licensed for application to existing data sets. Many commercial insurers are using these measures, although there is little evidence about the relative merits of competing products from different vendors. Because of concerns over the proprietary nature of these measures, some collaboratives, such as the Washington-Puget Sound Health Alliance, have elected instead to use public domain episode-based measures that are currently under development.

### *Further Methodological Questions*

Several methodological questions that are important to establishing credible resource use measurement remain. These questions apply to most types of measures.

1. **Reliability:** Reliability is an analysis of whether the variation seen in resource use is due to measurement error or to true differences in performance. The reliability of various resource use measures is largely unknown. The sample size of observations required to produce stable resource use estimates is uncertain. Health plans currently use arbitrary cutoffs, such as 30 episodes per physician, and therefore are often unable to profile as many as one-third of the eligible physicians in their networks.
2. **Provider Attribution:** A key issue for resource measurement for care provided by more than one provider, such as episodes of care, is how to attribute primary accountability for the resources used. Various algorithms, mainly based on visit counts and payment amounts, have been used. Different algorithms lead to different assignments, and every algorithm needs to be adjusted based on market characteristics such as the availability of subspecialists and geographic or cultural isolation. No national consensus guidelines for provider attribution are available.
3. **Risk adjustment:** Variation in resource use may be driven largely by differences in patient risk. While several risk-adjustment methods are used in various applications, most notably by vendors such as 3M Health Information Systems (distributor of Severity of Illness scores for All Patient Refined DRGs), limited testing has been done in some resource measurement applications, such as episodes of care.
4. **Treatment of outliers:** The distribution of resource use across individuals is highly skewed, with some people having no encounters or prescriptions and others having hundreds of encounters per year. Some users exclude outliers, but a preferable approach is probably to truncate (also known as Winsorizing) outliers to reduce their influence on subsequent analyses.

## Question 17. Which national groups are developing or endorsing resource use measures?

**The Agency for Healthcare Research and Quality** (AHRQ) includes a chapter on "efficiency" in its annual *National Healthcare Quality Report*.[115] The chapter includes several "potential" measures of efficiency that AHRQ believes "should be viewed as preliminary and designed to stimulate productive ongoing discussions about health care efficiency." These measures include trends in potentially avoidable hospitalizations and related costs, rehospitalization for heart failure, and an application of stochastic frontier analysis, which is an econometric technique that models provider-level inefficiency as a departure from an estimated best-practice frontier.[108,116] A set of resource use and efficiency measures is currently being developed for broader application, under AHRQ's Quality Indicators program.

**The Centers for Medicare & Medicaid Services** (CMS) is developing reporting of physician resource use using episode-based measures.[117] CMS has been evaluating two commercial episode groupers, Episode Treatment Groups (ETGs) and Medstat Episode Groups (MEGs). It also has funded a study developing alternative approaches to the commercial groupers. It is continuing to explore ways to improve on the commercial measures and is considering funding the development of new groupers for use with Medicare claims. CMS's publicly reported, National Quality Forum (NQF)-endorsed measures of 30-day readmissions after hospitalization for heart failure, pneumonia, or heart attack may also be interpreted as hospital-level resource use measures.

**The National Committee for Quality Assurance** (NCQA) has developed Relative Resource Use (RRU) measures. The RRUs are population-based measures that are used to compare health plans on resources used to care for beneficiaries with six conditions: asthma, diabetes, low back pain, cardiovascular disease, hypertension, and chronic obstructive pulmonary disease. Published tables allow organizations to match severity-adjusted resource use within service categories (Inpatient Facility, Surgery and Procedure, Evaluation and Management (E&M), and Pharmacy) to a standardized allowed payment in order to calculate total standard costs for their eligible members across different areas of clinical care.

**The Leapfrog Group** is a purchaser coalition that publishes information on hospital quality and resource use for coronary artery bypass graft surgery, percutaneous coronary intervention, heart attack, and pneumonia. Efficiency of care for each procedure and condition is a blend of a hospital's quality score for that procedure or condition (based on statewide outcome data, process-of-care measures, and volume) with their resource utilization score for that procedure or condition (Table 5). The resource utilization score for each procedure or condition is based on a hospital's standardized, risk-adjusted, geometric mean length of stay for that procedure or condition, inflated by the hospital's 14-day all-cause readmission rate for that condition.

**The Quality Alliance Steering Committee** (QASC) is a collaborative effort among government agencies, physicians, nurses, pharmacists, hospitals, health insurers, employers, consumers, and others.[117] To support the generation of effective health care performance information, the QASC is working to foster coordinated episode- and patient-level measures across the care continuum. It is aggregating data from different national health plans and Medicare to enable measurement of physicians' care for their entire practice panels. The QASC is also developing resource use measures for 20 high-cost/priority conditions, which will include both episode-based and per capita resource use measures.

**The National Quality Forum** (NQF) has developed a draft measurement framework for efficiency. A draft report, *Measurement Framework: Evaluating Efficiency Across Patient-Focused Episodes of Care*, has been endorsed by the NQF.[106] This episode-based framework will be used to develop a comprehensive set of performance measures, including resource use and quality measures, for selected clinical conditions.

**The Consumer-Purchaser Disclosure Project** (CPDP) is a multistakeholder collaboration involving consumer, employer, and labor organizations. It has published a "Patient Charter for Physician Performance Measurement, Reporting and Tiering Programs." This charter lays out a set of principles to guide physician performance measurement, which community quality collaboratives might consider as they develop reporting programs about resource use. The principles are[118]:

1. Measures should be meaningful to consumers and reflect a diverse array of physician clinical activities.
2. Those being measured should be actively involved.
3. Measures and methodology should be transparent and valid.
4. Measures should be based on national standards to the greatest extent possible.

## Question 18. How have resource use measures been used to compare providers to benchmarks?

**Provider resource use is typically compared to benchmarks of "peer" providers**. A common approach is to calculate a resource use score by dividing a provider's observed resource use by the "expected" resource use derived from a benchmark population. This approach allows for aggregation of resource use measurements across multiple episodes (or other units of service). The resource use score is then used to identify outliers that have significantly higher or lower resource use than peers.

Slightly different methods are used in these calculations, which could have a significant effect on the results.[119] For example, Medstat Episode Group (MEG) can be used to calculate a Risk-Adjusted Cost Index (RACI), which is a ratio of the total allowed costs of qualified episodes for which the provider is attributed responsibility divided by the total expected costs. The expected cost is based on the average cost of similar episodes based on MEGs, severity of illness (classified from 0 to 3), comorbidity burden (defined by the Diagnostic Cost Group [DCG] Relative Risk Score), provider specialty, and geographic region. With Symmetry Episode Treatment Groups (ETGs), the estimated ratio is somewhat different: a physician's normalized, actual resource use for a given set of ETGs serves as the numerator, and his or her specialty's normalized, average resource use for the same set of ETGs serves as the denominator.

Arbitrary thresholds are typically used to determine which providers are categorized as "high resource use" or "low resource use." These **thresholds are often set at percentiles in the distribution of resource use scores** (e.g., the providers in the top decile of resource use scores are labeled "high resource use"). The **Washington-Puget Sound Health Alliance** provides another example of how resource use is compared. Following the format of their quality reports, the Alliance plans to report provider resource use for episodes of care as “above the regional average,” “at the regional average,” or “below the regional average.”

### *Choosing Benchmarks*

Different peer groups have been used in resource use comparisons. One decision is **whether to use a benchmark of providers from the same geographic area or a national benchmark**. Practice patterns vary widely between regions. For example, the cost per episode and the number of episodes per beneficiary were found to differ widely between Minneapolis and Miami.[120] A regional benchmark would control for these differences, while a national benchmark would compare providers to national practice standards.

A second decision is **whether to use a benchmark of providers of the same type (e.g., specialty) or providers of all types (e.g., multiple specialties)**. Many conditions are treated by multiple specialties, and practice patterns often differ widely by specialty. For example, endocrinologists and primary care physicians both provide care for diabetes. A measure of resource use for diabetes treatment could compare endocrinologists to other endocrinologists, as well as to primary care physicians.

A single-specialty benchmark holds providers accountable for the standards of their specialty, while a multiple-specialty benchmark compares resource use across specialties. Similarly, hospital comparisons could be limited to hospitals of the same teaching status or safety-net status. Comparing providers only to similar providers has the advantage of reducing variation that is beyond the provider’s control, but the disadvantage is that providers are not held accountable to the level of performance achieved by other types of providers caring for similar patients.

# Part IV. Selecting Quality and Resource Use Measures

The first three sections of this Decision Guide provide background information for community quality collaboratives to consider before selecting quality and resource use measures. In this section, Questions 19-25 provide a framework that collaboratives may use to organize the information available to them. They also can use this framework to design a measure selection process that is consistent with accepted theoretical constructs as well as the collaborative's own stated objectives and constraints. We start with a brief overview of national initiatives related to the standardization of measures, to provide context for this topic.

## Question 19. What national initiatives and forces are driving the standardization of quality measurement?

Historically, the **National Committee on Quality Assurance** and **The Joint Commission** initiated the standardization of quality measurement. Increasingly, the momentum is being carried forward by providers' desire for harmonized measures to minimize the burden of data collection and organizations' desire for more comparative national and regional benchmarks. As stakeholders in a community adopt the same measures, they conserve resources that would otherwise be diverted to developing or testing their own measures, and they send a stronger "market signal" to providers to improve quality in those specific areas.

In concert with the standardization movement is the health information technology (HIT) effort that is rapidly progressing with the coalescing of public (Federal) and private support (perhaps most notable is the $19 billion HIT funding earmarked in the 2009 Federal stimulus package).[121] Multiple national initiatives are trying to build on this effort to establish broadly supported, valid measures of quality and cost that can be used nationwide. Community quality collaboratives may consider participating in, or at least monitoring, these national initiatives that will influence how quality and resource use measures are collected and reported over the next decade.

### *National Initiatives for Standardizing Measurement*

The **National Priorities Partnership (NPP)** is a collaborative effort of 28 major national organizations, convened by the National Quality Forum (NQF) to represent multiple stakeholders. These stakeholders include consumers, employers, government, health plans, health care organizations and professionals, scientists, accrediting and certifying bodies, and quality alliances. As a first step, the Partners "identified a set of six National Priorities and Goals to help focus performance improvement efforts on high-leverage areas—those with the most potential in the near term to result in substantial improvements in health and health care—and thus accelerate fundamental change in our health care delivery system."[122] As a second step, the Partners have agreed to work together over the next year to align the drivers of change, such as payment reform, accreditation and certification, and performance measurement, around these goals:

1. Engaging patients and families in managing their health and making decisions about their care.
2. Improving the health of the population by creating communities that foster health and wellness.
3. Improving the safety and reliability of America's health care system.

4. Ensuring that patients receive well-coordinated care within and across all health care organizations, settings, and levels of care.
5. Guaranteeing appropriate and compassionate care for patients with life-limiting illnesses.
6. Eliminating overuse while ensuring the delivery of appropriate care, continually and safely reducing the burden of unscientific, inappropriate, and excessive care.

To implement the vision of the NPP, a stronger national infrastructure is needed for health care performance measurement and reporting. The **Quality Alliance Steering Committee (QASC)** was formed in 2006 as a collaborative effort to ensure that quality measures are constructed and reported in a clear and consistent way that informs both consumer and employer decisionmaking, as well as the efforts of practitioners to improve. QASC includes existing quality alliances, government agencies, physicians, nurses, hospitals, health insurers, consumers, accrediting agencies, and foundations. Convened by the Brookings Institution with financial support from the Robert Wood Johnson Foundation, the QASC's High-Value Health Care Project is testing approaches to combining data from the public and private sectors to measure and report on physician practices in a meaningful and transparent way for consumers and purchasers of health care. Based on these pilot projects, the QASC hopes to develop an infrastructure for combining summary provider information from Medicare and private health plans at the national level, offering a more complete picture of providers' cost and quality.

An overlapping group of nearly 130 organizations has recently joined together as **Stand for Quality in Health Care** (www.standforquality.org) to develop and support specific recommendations for policymakers considering health care reform. In a report titled *Building a Foundation for High Quality, Affordable Health Care: Linking Performance Measurement to Health Reform*, these organizations argue for dedicated Federal support for six key functions of the performance measurement, reporting, and improvement enterprise:

1. Set national priorities through a multistakeholder process (based on the NPP) and provide ongoing coordination and self-evaluation.
2. Endorse and maintain valid, reliable, evidence-based, feasible, actionable, and usable measures for national use through a multistakeholder process (based on the NQF).
3. Develop measures to fill gaps in priority areas, including care coordination and transitions; palliative and end-of-life care; overuse and waste; promotion and adoption of healthy lifestyles; episode-based outcomes, processes, and costs; and disparities.
4. Implement effective and open consultative processes so that stakeholders can inform policies on use of measures (based on the alliances described further below).
5. Collect, analyze, and make performance information from health plans, clinicians, nursing homes, hospitals, clinics, and other providers available and actionable at the local, State, and national levels (based on the QASC).
6. Support a sustainable infrastructure for quality improvement in all settings.

AHRQ's Chartered Value Exchange (CVE) program, formed in 2007, brings together 24 CVEs, or community quality collaboratives, from across the country. In aggregate, these collaboratives involve more than 575 health care leaders and represent more than 124 million lives, which is more than one-third of the U.S. population. The collaboratives are multistakeholder initiatives with a mission of quality improvement and transparency. The program is built on three overarching principles:

1. All health care is “local.” National goals and common standards are important, but real improvement needs to take place in local settings where the various stakeholders know and work with one another.
2. Transparency in measuring and reporting accurate and meaningful information on quality and cost is key to helping providers improve and consumers become engaged managers of their own health and health care.
3. Collaboratives involving key stakeholder groups (e.g., public and private payers, providers, plans, consumer organizations, State data organizations, quality improvement organizations, health information exchanges) hold the promise to foster requisite reforms.

Through AHRQ’s national Learning Network, CVE members learn from each other and from experts, sharing experiences and best practices through meetings, Web conferences, documents, and an electronic bulletin board. The Learning Network’s areas of focus are driven by the needs of the CVEs and include: collaborative leadership and sustainability; public at-large engagement; quality and efficiency measurement; public reporting; provider incentives; consumer incentives; coordinated cross-organizational, cross-stakeholder quality improvement; and health information technology/health information exchange. AHRQ tools for CVEs and other community collaboratives are available at http://www.ahrq.gov/qual/value/localnetworks.htm.

To facilitate planning and coordination at both the national level and in CVE communities, AHRQ regularly meets with leaders of other key community-based quality improvement initiatives. Some of these organizations include QASC and its National-Regional Implementation Workgroup; the Robert Wood Johnson Foundation Aligning Forces for Quality Program; and the Network for Regional Healthcare Improvement (NRHI).

**Aligning Forces for Quality (AF4Q)** (www.rwjf.org/qualityequality/af4q/index.jsp) helps participating communities work toward sustainable health care quality through the leadership of a multistakeholder local alliance. These alliances focus on three key intersecting program areas: (1) developing a local quality improvement resource that will help health professionals improve once they recognize the need, (2) helping the public in those communities get substantially better at using appropriate information in making health and health care decisions, and (3) working to substantially increase performance measurement and public reporting of those measures. AF4Q communities actively promote nurse leadership in the effort and also focus on racial and ethnic disparities in all aspects of the initiative. By aligning key stakeholders to address these and other key issues, AF4Q aims to achieve communitywide, sustainable transformation of health care. Teams in AF4Q communities improve the quality of information about physician performance and public access to that information.

**There is tremendous synergy between the CVE and AF4Q initiatives**; 13 of the 15 AF4Q sites are also affiliated with CVEs. For example, the Health Improvement Collaborative of Greater Cincinnati simultaneously functions as a CVE and leads the Cincinnati AF4Q (www.rwjf.org/qualityequality/af4q/communities/cincinnati.jsp). It is working to enhance local infrastructure and to align key drivers of overall health care improvement through a condition-specific approach. Currently, they are focused on diabetes-care messaging among employers, health plans, providers, and community-based organizations; initiating quality improvement among primary care providers; and initiating regionwide public reporting of selected primary care practices' quality-related outcomes.[123] A collaboration in Wisconsin (www.rwjf.org/qualityequality/af4q/communities/wisconsin.jsp) provides another example of the confluence of both initiatives.

### *Hospital-Related Initiatives*

The **Hospital Quality Alliance (HQA)**, a multistakeholder alliance concerned with hospital quality of care, has taken a leading role in selecting hospital quality measures for presentation on the Hospital Compare Web site developed by the Centers for Medicare & Medicaid Services (CMS).[124] Its role as an adopter of measures will be discussed further under Question 21.

In the aftermath of several studies demonstrating the advantages of enhanced administrative data,[26,93] AHRQ is sponsoring **"Adding Clinical Data Pilot and Planning Projects"** to link clinical data, especially present on admission coding and laboratory results, to existing administrative data sets collected for the Healthcare Cost & Utilization Project (HCUP). This method is seen as a practical, effective, and cost-effective way to produce more accurate and thorough quality assessments of hospitals. Results from the pilot sites in Florida, Minnesota, Virginia, and Washington were expected in late 2009 (www.hcup-us.ahrq.gov/reports/clinicaldata.jsp).

### *Physician-Related Initiatives*

The **AQA Alliance** (www.aqaalliance.org) was founded by physician organizations (in collaboration with AHRQ and America's Health Insurance Plans) as the Ambulatory Care Quality Alliance but has expanded to include other stakeholders interested in measuring the quality of care provided by physicians and other licensed professionals. Similar to HQA, the approval of measures by the AQA has been a factor considered by CMS in the selection of physician quality measures for use in public reporting and pay-for-performance programs. Its role as an approver of measures is discussed further under Question 21.

**Bridges to Excellence (BTE)** is a national pay-for-performance program with a standard data exchange platform and performance measurements (http://bridgestoexcellence.org). This program is currently available in limited markets across the United States and is closely aligned with NCQA efforts. Physicians self-report common performance measures endorsed or approved by NQF, AQA, or the American Medical Association's Physician Consortium for Performance Improvement (PCPI) in nine clinical areas. Physician-specific performance results are not publicly available; however, participants use the data for quality improvement and pay-for-performance programs. BTE-recognized physicians are listed on the HealthGrades® Web site.

Before physician performance measurement can reach its full potential nationally, the market needs strong consensus standards for managing and auditing the measurement process, which is led by health plans and other organizations. **The Consumer-Purchaser Disclosure Project (CPDP)** created a "Patient Charter for Physician Performance Measurement, Reporting and Tiering Programs." The charter establishes a national set of principles to guide measuring and reporting to consumers about doctors' performance. The CPDP intends for health plans and others to adopt the Patient Charter and abide by CPDP's guidelines for physician measurement. The guidelines include auditing the measurement process imposed by health plans to reduce administrative burden and to ensure measurement transparency for participating physicians. NCQA was named as the first approved independent reviewer to certify organizational compliance with these guidelines.

Details about the national initiatives regarding resource use measure development and implementation can be found in Question 17 of this Decision Guide.

## Question 20. How can the Institute of Medicine's six "quality domains," the National Priorities Partnership's six National Priorities, and Donabedian's "structure, process, and outcome" typology be used to select appropriate measures of quality?

### *Institute of Medicine Framework*

The Institute of Medicine's (IOM) 2001 report, *Crossing the Quality Chasm*, identified six "aims for improvement to address key dimensions in which today's health care system functions at far lower levels than it can and should."[3] These quality domains include:

1. **Safety –** avoiding injuries to patients from the care that is intended to help them;
2. **Effectiveness** – providing services based on scientific knowledge to all who could benefit and refraining from providing services to those not likely to benefit;
3. **Patient centeredness –** providing care that is respectful and responsive to individual patient preferences, needs, and values;
4. **Timeliness –** reducing waits and sometimes harmful delays for both those who receive and those who give care;
5. **Efficiency** – avoiding waste, including waste of equipment, supplies, ideas, and energy; and
6. **Equity –** providing care that does not vary in quality because of personal characteristics such as gender, ethnicity, geographic location, and socioeconomic status.

This framework gives community quality collaboratives a useful way to conceptualize where they want to go as they move forward to improve health care. It also helps collaboratives to think about what they want to measure and whether the available set of measures adequately covers the domains of concern. In general, **the largest number of measures available to collaboratives is in the domain of effectiveness**, but the **CAHPS® (Consumer Assessment of Healthcare Providers and Systems) family of surveys has greatly expanded opportunities for measuring both patient centeredness and timeliness**. In the last few years, there has been an explosion of efforts to measure quality in the previously neglected domains of safety, efficiency, and equity.

### *National Priorities and Goals*

Under the auspices of the National Quality Forum (NQF), the National Priorities Partnership (NPP) (described under Question 20) has identified a set of six national priorities and goals to help focus performance measurement and improvement on "high-leverage areas." These areas have "the most potential to result in substantial improvements in health and health care" and "thus accelerate fundamental change in our healthcare delivery system." These goals complement and update the IOM framework, as follows.

1. **Patient-centeredness:** Engaging patients and families in managing their health and making decisions about their care. Health care should honor "each individual patient and family, offering voice, control, choice, skills in self-care, and total transparency," adapting readily to individual and family circumstances, and to differing cultures, languages, and social backgrounds.
2. **Equity and population health:** National, State, and local systems of care should be fully invested in the prevention of disease, injury, and disability, "helping all people reduce the risk and burden of disease."
3. **Safety and effectiveness:** America's health care system should be "relentless in continually reducing the risks of injury from care, aiming for zero harm wherever and whenever possible," "guaranteeing that every patient, every time, receives the benefits of care based solidly in science."
4. **Timeliness and coordination of care:** Ensuring that patients receive well-coordinated care within and across all health care organizations, settings, and levels of care. "A healthcare system should guide patients and families through their healthcare experience, while respecting patient choice, offering physical and psychological supports, and encouraging strong relationships between patients and the professionals accountable for their care."
5. **Appropriate and compassionate care for patients with life-limiting illnesses:** Health care should promise "dignity, comfort, companionship, and spiritual support to patients and families facing advanced illness or dying, fully in synchrony with all of the resources that community, friends, and family can bring to bear."
6. **Efficiency:** Eliminating overuse while ensuring the delivery of appropriate care. Health care should promote "better health and more affordable care by continually and safely reducing the burden of unscientific, inappropriate, and excessive care, including tests, drugs, procedures, visits, and hospital stays."

### *Donabedian Typology*

Another useful typology of quality measures was developed by Avedis Donabedian about 30 years ago,[125-127] based on earlier work by Sheps in the 1950s. Donabedian described three approaches to acquiring information about health care quality:

1. **Structural measures focus on the conditions under which care is provided.** These include the material (e.g., facilities, equipment) and human resources (e.g., staffing ratios, qualifications, experience) available to provide care, as well as the organizational context (e.g., size, volume, IT systems) that facilitates or impedes the delivery of optimal care.
2. **Process measures focus on what a health care provider does to maintain or improve patients' health**, including appropriate and evidence-based screening, diagnosis, treatment, rehabilitation, education, and prevention.
3. **Outcome measures focus on changes in health status that are attributable to health care**, including mortality, morbidity (e.g., complications, unplanned readmissions), functional status, quality of life, and health-related knowledge and behaviors.

### *IOM-Donabedian Matrix*

Although the IOM framework and the Donabedian framework categorize quality measures differently, one can create **a two-dimensional matrix to help clarify how well one is covering all of the domains of interest** (Table 7). Within each IOM domain, measures of structure, process, and outcome are typically available. Two examples of structural measures would include quality improvement systems and adequate nurse staffing, both intended to facilitate safe and effective care.

Some structures facilitate equitable care, such as adequate interpreting services. There may also be structures that facilitate patient-centered care, such as providers' use of patient survey data to improve patient-centered care. Structures facilitating timely care might include health maintenance organization (HMO) policies on prior authorization and provider policies on scheduling urgent care appointments.

Process and outcome measures can address almost any IOM domain of care, although there is clearly some overlap across domains. For example, among the AHRQ Quality Indicators, the Patient Safety Indicators may serve as indicators of safe inpatient care, the Prevention Quality Indicators as indicators of timely and effective outpatient care, the mortality-based Inpatient Quality Indicators or IQIs as indicators of effectiveness, and utilization-based IQIs as indicators of resource use.[74]

**Table 7. Matrix of quality measure typologies with examples**

| IOM Domains | Structure | Process | Outcome |
|---|---|---|---|
| **Effective** | Cardiac nurse staffing, nursing skill mix (RN/total) | Use of angiotensin-converting enzyme (ACE) inhibitor or angiotensin receptor blocker (ARB) for patients with systolic heart failure | 30-day readmissions (or mortality) for heart failure |
| **Patient Centered** | Use of survey data to improve patient-centered care | Did the nurses treat you with courtesy and respect? | Overall rating of care |
| **Timely** | Physician organization policy on scheduling urgent appointments | Received beta blocker at discharge and for 6 months after AMI | Potentially avoidable hospitalizations for angina (without procedure) |
| **Safe** | Computerized physician order entry with medication error detection | Use of prophylaxis for venous thromboembolism in appropriate patients | Postoperative deep vein thrombosis or pulmonary embolism |
| **Efficient** | Availability of rapid antigen testing for sore throat | Inappropriate use of antibiotics for sore throat | Dollars per episode of sore throat |
| **Equitable** | Availability of adequate interpreting services | Use of interpreting services when appropriate | Disparity in any other outcome according to primary language |

To achieve our national goals for effective, patient-centered, timely, safe, efficient, and equitable care, as set forth by the IOM and recently reinforced by the NPP, **quality measurement will eventually need to cover all domains and borrow from all approaches**. However, few markets will start with the expertise to collect and report measures in all domains using multiple approaches. Community quality collaboratives may want to consider prioritizing and developing **a multiphase strategy for using available data and measures**. This matrix may be helpful in facilitating such a phased approach, which begins by capturing “low hanging fruit” and then gradually expands to cover a broad range of domains and approaches.

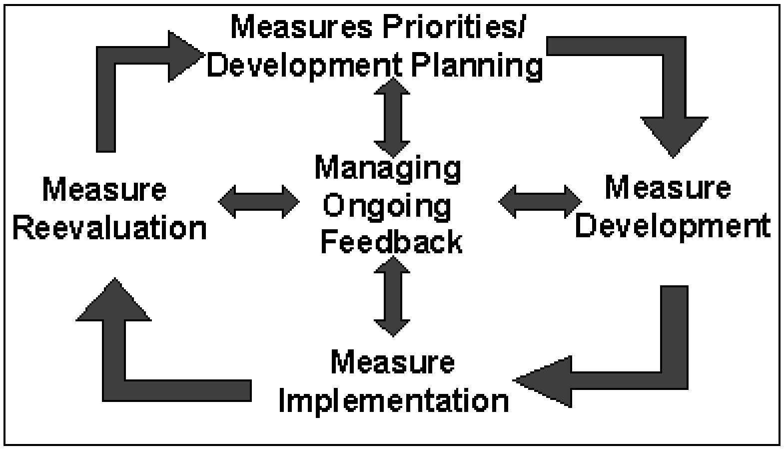
Although the greatest number of available measures is still in the effectiveness domain (using the process approach), stakeholders are increasingly seeking out a broader range of measures in other IOM domains, including outcome measures. As described at AHRQ’s Talking Quality Web site (www.talkingquality.gov), providing consumers with a clear framework for understanding quality also helps them to grasp the value and relevance of a broader range of quality indicators.

## Question 21. What are the roles and responsibilities of the organizations that endorse or approve measures versus those organizations that develop measures?

### *Measure Developers*

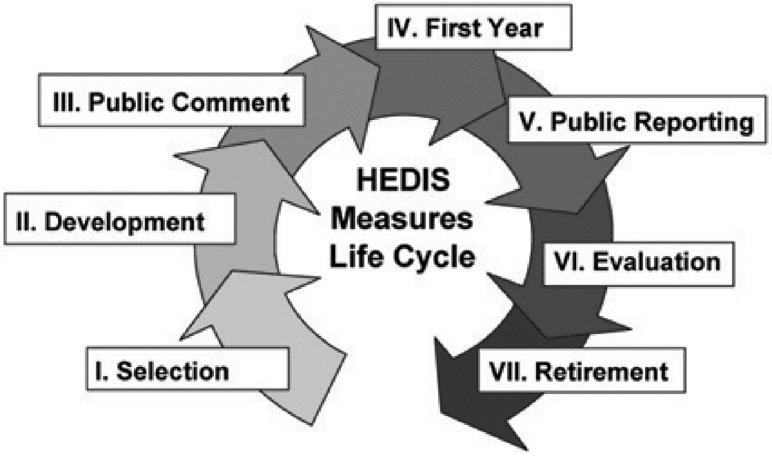
Measure developers use a standardized approach to creating, maintaining, and retiring quality performance indicators. After prioritizing medical conditions or interventions of interest, they identify and recruit experts to research evidence-based literature to identify candidate measures. Potential process measures are often based on clinical practice guidelines developed by professional organizations and evidence reviews supported by AHRQ or similar organizations. Technical specifications for data collection and measure calculation are field tested for validity, reliability, and feasibility. Figure 1[128] and Figure 2[129] illustrate the Centers for Medicare & Medicaid Services (CMS) and National Committee for Quality Assurance (NCQA) approaches to measure development.

**Figure 1: Overview of CMS quality measures development process**



**Source:** CMS Measures Management System Blueprint, version 7.0. The Blueprint is publicly available on the Quality Measures Management Information section of the CMS Web site (**https://www.cms.hhs.gov/apps/QMIS/default.asp**).

**Figure 2: NCQA Healthcare Effectiveness Data and Information Set (HEDIS) measures lifecycle**



**Source:** NCQA, 2009.

Measure developers include a number of professional associations, accrediting bodies, and government entities at the national level. Among them are the American College of Cardiology, AHRQ, American Cancer Society, American Medical Association and its affiliated organizations (structured as the Physicians' Consortium for Performance Improvement), CMS, The Joint Commission, and NCQA. These organizations submit measures for review by endorsement and approval organizations.

### *Measure Endorsers, Approvers, and Adopters*

Measure endorsers or "approvers" use a consensus-based approach to evaluate the feasibility, reliability, validity, and usability of quality performance measures. The process is similar to the measure development process and includes multiple stakeholders to reflect varied perspectives and expertise. The **National Quality Forum (NQF)** is the premier organization for evaluating health care performance measures throughout the continuum of care.[130] It uses a consensus development process, as shown in Figure 3,[131] to assess and endorse, when appropriate, voluntary measures submitted by various measure developers. A list of NQF's currently endorsed measures is available at www.qualityforum.org. The criteria that NQF uses in evaluating candidate standards are further described in Question 22.

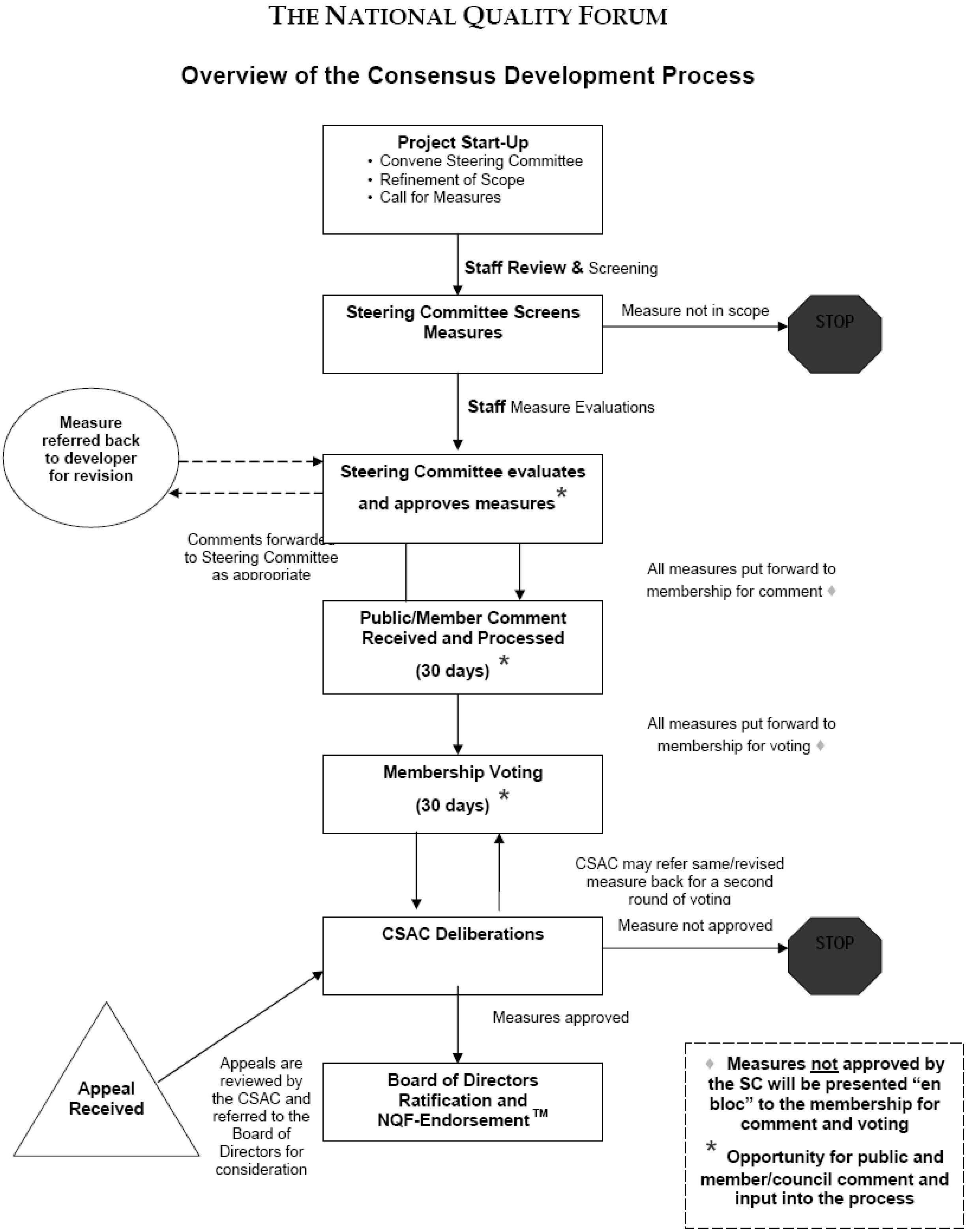
The **Hospital Quality Alliance (HQA)**, a multistakeholder alliance concerned with hospital quality of care, characterizes itself as a "measure adopter" and selects hospital measures from among those previously endorsed by NQF.[124] By implementing consensus-based, nationally standardized performance measures, HQA promotes a common, unified approach to measurement and reporting. Its current compendium of adopted measures is at www.hospitalqualityalliance.org/hospitalqualityalliance/qualitymeasures/qualitymeasures.html. Because HQA-adopted measures feed into CMS's Hospital Compare Web site, community quality collaboratives often rely on HQA recommendations when selecting measures for their own reporting programs. However, CMS has stated that Federal law only indicates "that measures must reflect consensus among affected parties and, to the extent feasible and practicable, must include measures set forth by one or more national consensus building entities… the Secretary is not required to limit measures to those endorsed or adopted by any particular consensus organization or quality alliance…"[132]

The **AQA Alliance** (www.aqaalliance.org ) is a voluntary, multistakeholder collaborative of physicians and other clinicians, consumers, purchasers, health plans, and other interested parties. They have joined together to determine how to most effectively and efficiently improve performance measurement at the clinician or group level. In addition, they examine ways to collect and aggregate performance data and report meaningful information to consumers, clinicians, and other stakeholders to inform decisionmaking and improve outcomes. The AQA has published specific criteria for approving measures. The criteria emphasize the value of measure sets that are aligned with the IOM's priority areas, evidence based, complementary of hospital/facility measures, and focused on high-impact problems. The AQA Alliance defers to NQF as the final arbiter for measure endorsement ("for indicators not endorsed by the NQF during their call for measures and endorsement process, AQA-approval will be rescinded").[133] Its current compendium of approved measures is at www.aqaalliance.org/files/ CompendiumofApprovedMeasures.doc.

The **American Medical Association Physician Consortium for Performance Improvement (PCPI)** offers the opportunity for professional collaboration on developing, testing, and maintaining evidence-based clinical performance measures for physicians. It also reviews measures developed independently.[134] PCPI's role is self-limited to development and adoption rather than endorsement; PCPI measures are submitted to AQA and NQF for approval and endorsement. **The Joint Commission** and **NCQA** are also measure developers that sometimes adopt measures developed by other organizations.

The Joint Commission's Core Measures are used to inform the hospital accreditation process and NCQA's HEDIS measures are used to inform the plan accreditation process. However, these measures generally are not used by community quality collaboratives for public reporting or pay for performance until they are approved or endorsed by one or more of the three organizations described above (i.e., NQF, HQA, AQA).[135] In this way, collaboratives can avert potential criticism for reporting "experimental" measures that are not accepted by stakeholders at the national level. Some collaboratives may implement unendorsed measures, at least in confidential reporting, as a means of testing them, demonstrating their usefulness and feasibility, and thereby helping to move them through the endorsement process.

**Figure 3: The NQF consensus development process**



**Source:** NQF, 2008.
SC=Steering Committee; CSAC=Consensus Standards Approval Committee

## Question 22. What criteria should we use when screening measures of quality for public reporting or other purposes?

### *Screening Framework*

Once a community quality collaborative establishes its purposes in assessing health care quality (e.g., pay for performance, public reporting, internal quality improvement), it must screen a large number of established quality measures. This process can be daunting, but community quality collaboratives can take advantage of **several evidence-based and consensus-based evaluation frameworks** that leading organizations have developed to prioritize measures.[107,131,136] Although these evaluation frameworks were developed for somewhat different purposes and were initially applied to different sets of candidate indicators, they are actually quite similar and are therefore useful to collaboratives that hold a variety of measurement agendas.

As shown in Appendix A, the National Quality Forum (NQF) built on earlier work by The Joint Commission, National Committee for Quality Assurance (NCQA), and Institute of Medicine (IOM) (on behalf of AHRQ's National Healthcare Quality Report [NHQR][137]) to propose a four-domain scheme for evaluating quality measures. Within each of these domains are several key questions or criteria, as described below. Depending on local priorities, a community quality collaborative may put more weight on one domain and less weight on others, and it may choose to focus on a single criterion or set of criteria within each domain. In some cases**, a collaborative may accept NQF endorsement of an indicator as both necessary and sufficient evidence that the indicator is acceptable for public reporting**. In other cases, **a collaborative may set a lower or higher threshold than NQF**, perhaps because of differing local views about the importance of a quality-related problem, local availability of better or worse data, or local interest in testing a measure that may be submitted for endorsement in the future.

Here is a brief summary[138] of **NQF's four domains for evaluation**:

| | |
|---|---|
| **Importance:** | There should be a leverage point for improving quality, considerable variation in quality of care, or suboptimal performance in the area of interest. |
| **Scientific acceptability:** | The measure should be well defined, precisely specified, reliable, and able to produce the same results a high proportion of time in the same population. It should be valid, accurately representing the concept being evaluated. The measure should be precise, adequately discriminating between real differences in provider performance and adaptable to patient preferences and a variety of settings. An adequate and specified risk-adjustment strategy should be available and there should be evidence linking process measures to outcomes. |
| **Usability:** | The measure can be used for making decisions and implementing change. Performance differences should be statistically meaningful (practically and clinically). The measure should provide for appropriate risk stratification, risk adjustment, and other forms of recommended analyses. |
| **Feasibility:** | Data collection should be linked to care delivery when feasible, and timing and frequency of measure collection must be specified. The benefit of implementation should be evaluated against financial and administrative burden. Confidentiality concerns should be addressed and an audit strategy should be available. |

Scientific acceptability, which aligns with the NCQA and IOM/NHQR domains of "scientific soundness," may be the most difficult domain for community quality collaboratives to assess, because it draws on complex concepts such as reliability, validity, and lack of bias. Many chartered value exchanges (CVEs) lack the technical expertise and resources to assess reliability, validity, and bias, but they should be intelligent consumers of information about these performance characteristics from other sources (such as measure developers). The following questions from the National Quality Measures Clearinghouse, which measure developers/sponsors are expected to answer, are useful for understanding measure validity[139]:

**Q1.** How strong is the scientific evidence supporting the validity of this measure as a quality measure?
**Q2.** Are all individuals in the denominator equally eligible for inclusion in the numerator?
**Q3.** Is the measure result under control of those whom the measure evaluates?
**Q4.** How well do the measure specifications capture the event that is the subject of the measure?
**Q5.** Does the measure provide for fair comparisons of the performance of providers, facilities, health plans, or geographic areas?
**Q6.** How well does your intended use of the measure match the developer's intended use?
**Q7.** Does the quality of data available meet the measure standards (e.g., reliability, appropriate sample size, accessibility)?

### *Screening Process*

Three comprehensive databases are available to community quality collaboratives to help them identify and screen potential measures.[133] The **National Quality Measures Clearinghouse** (NQMC), supported by AHRQ, summarizes measures that are submitted to it by many different measure sponsors. These include government agencies, accrediting bodies, research institutions, professional societies, and even individual hospitals and health systems. The NQMC provides detailed information about each measure, based on a template of measure attributes that includes each measure's:

- Title and source
- Domain of measurement (e.g., structure, process, outcome)
- Description and rationale
- Supporting evidence of value
- Current use
- Care setting
- Professionals responsible
- Target population age and gender
- Incidence or prevalence
- Association with vulnerable populations
- Burden of illness
- Associated utilization and costs
- IOM domain
- Sampling frame
- Denominator and numerator definitions and time windows

- Data source
- Scoring
- Risk adjustment
- Standards of comparison (e.g., benchmarks)
- Evidence of reliability or validity
- Endorsements

An online search utility allows users to find the subset of measures that meet specific criteria, based on this template of attributes, while a measure comparison utility allows users to generate side-by-side comparisons for any combination of two or more measures. Links to full-text measures and ordering details are provided when available.

The NQMC also provides a **"Measure Archive"** where withdrawn and revised measures are housed. At the time this Decision Guide was written, 437 measures had been updated or withdrawn by measure developers. This large number illustrates the constant changes in the measurement field; therefore, users need to be diligent in monitoring changes to nationally accepted measures. For example, changes to numerator or denominator definitions may limit a CVE's ability to compare measures year to year and may necessitate modification of predefined pay-for-performance methodologies. In addition to suggesting that users contact measure developers directly for the most current information, the NQMC provides an e-mail service that alerts subscribers to new information (www.qualitymeasures.ahrq.gov/whatsnew/ subscription.aspx).

The **National Quality Forum** (NQF) maintains an ongoing list of approved measures, which it refers to as **National Voluntary Consensus Standards**. The list currently available at the NQF Web site (www.qualityforum.org/Measures_List.aspx) includes 545 standards covering all domains of inpatient and outpatient care. Each standard is assigned an official number and title, succinctly described, attributed to a specific "steward," and classified in terms of the following:

- Care setting (e.g., ambulatory care, hospital, home health, hospice, nursing home, dialysis center)
- Type (e.g., structure, process, outcome, patient experience)
- Level of measurement (e.g., facility, individual clinician)
- Data source (e.g., paper medical record, electronic claims, clinical registry)
- Endorsement status and date

This NQF list can be searched to identify, for example, the subset of endorsed measures related to hospital readmissions. Note that some of the listed information was not available online at the time of this publication.

The third resource, the **Centers for Medicare & Medicaid Services (CMS) Quality Measures Management Information System** (QMIS), is a comprehensive, Web-based repository of quality measures used by all CMS health care quality initiatives (https://www.cms.hhs.gov/apps/QMIS/default.asp). It was designed as an electronic tool to support the CMS Measures Management System. This system is a set of processes and decision criteria used to oversee the development, implementation, and maintenance of health care quality measures throughout their life cycle. The QMIS serves as the authoritative repository of information on the quality measures used by CMS, including their technical specifications, justification, and history. It is also being used to track the development of new measures and the maintenance of existing measures, providing a consistent mechanism by which requests, inquiries, and comments pertaining to measures can be processed. Users can browse or search for measures by name, description, approval status, clinical condition, developer, and contractor.

### *Community Collaborative Example*

The Alliance, a purchaser coalition in Wisconsin, uses a **two-phase process** to reduce the number of measures to a manageable number and it completely reevaluates the measure set every other year. The Alliance included hospitals at the beginning of the evaluation to help determine the measure set and circulated a list of almost 200 measures for comment. This process proved to be very cumbersome and was subsequently revised to include a shorter list of the most viable measures. The Alliance solicited feedback on the truncated list and considered alternative measures as suggested by hospitals. This new process was less overwhelming to the hospitals and resulted in what Alliance leaders consider to be a strong measure set.[140]

The Alliance also created a matrix that identifies the measure criteria that relate to their pay-for-performance program and those criteria that relate to public reporting. Most of the criteria apply to both goals, but some only apply to a single goal. For instance, measures that demonstrate variation in care are more important to public reporting than to pay for performance because provider reward is based on meeting certain performance standards.

## Question 23. Against which benchmarks should we measure our local performance?

The benchmark concept, originated by the manufacturing industry, is a critical tool for health care quality performance measurement that provides context for measuring individual performance. **A benchmark is a reference point or standard against which individual performance can be assessed.** Keife, et al., state that benchmarks should reflect the best care achieved for at least 10% of the eligible patient population. This standard means that a benchmark will always surpass average performance and should represent an attainable (clinically realistic) level of excellence. This approach has been described as **"Achievable Benchmarks of Care"** (ABC).[141,142]

Examples of benchmark use can be found through QualityCheck, sponsored by The Joint Commission, and CMS Hospital Compare, both of which offer the national average and a "benchmark" representing the top 10% of hospitals (also known as the 90th percentile) reporting that measure. Wessell, et al. (2008), demonstrated the feasibility of applying this ABC method in primary care settings through the Practice Partner Research Network, which includes 87 EMR-equipped practices with 712,000 patients across 35 States.[143]

In practice, however, benchmarks are often based on **average or median performance scores**, especially for risk-adjusted outcome comparisons such as those published by New Jersey, Pennsylvania, California, and many other States. Average benchmarks may be more palatable to the organizations being evaluated and easier to incorporate into statistical analyses, given the so-called "null hypothesis" that all organizations perform at the same, average level.

### Local/Regional Versus National Benchmarks

The question of whether to use benchmarks derived from local, State, regional, or national data depends on several factors, including the availability of benchmark data, the collaborative's objectives, and local performance levels.

**The availability of reliable benchmarks is the foremost consideration**. National benchmarks for measure sets developed by national organizations (e.g., National Committee for Quality Assurance's [NCQA's] Healthcare Effectiveness Data and Information Set, The Joint Commission's QualityCheck, AHRQ's HCUPNet utility and annual *National Healthcare Quality Report* and *State Snapshots*) are generally easy to obtain and offer reliable and valid comparisons to the Nation's performance. However, benchmarks for many physician group performance measures are not yet available at the national level. Instead, local and regional initiatives are establishing their own benchmarks, such as for coronary artery bypass surgery mortality in Pennsylvania, New York, California, and a few other States.

**Benchmarks at the local or State level may also be required if a collaborative creates unique performance measures or if local population factors (e.g., a shortage of primary care physicians) have a substantial impact on performance** at the local or State level. Although local benchmarks reflect local practice, they have an important disadvantage in that they are susceptible to undesirable local variation in the quality and pattern of care.[144] If an area compares poorly with other areas nationally, no real benefit from a local benchmark will be realized; scores will remain lower than what could be attainable. Other challenges for local benchmark creation and use include accruing a sufficiently large sample and addressing possible financial and political barriers that may hinder the creation of locally based benchmarks.

Another factor influencing the choice of benchmark is the **reason for measurement (pay for performance, internal quality improvement, or public reporting) in the context of current local performance levels**. For example, organizations using pay for performance as an incentive for improved quality may require providers to achieve a national benchmark, also known as a threshold, for supplemental payment. Using a national benchmark that is already vetted should allay provider concerns about its reliability and validity.

Dudley, et al., explore using either relative (local) or absolute performance thresholds to inform pay-for-performance programs.[145] Either is useful depending on the *primary* program goal: to improve the quality of care delivered by all eligible providers or to reward the highest quality providers.[145] When publicly reporting performance results, it is useful for consumers to put the information into a larger context. For practices that are high-volume Medicaid providers, such as Federally Qualified Health Centers, using NCQA clinical benchmarks that are specific to Medicaid may be perceived as being fairer than the alternatives.

### Peer Group Benchmarking

Another important question is whether to compare each provider organization to all of the other provider organizations in the market, or only to those with similar structural characteristics. The latter approach is also known as **peer group benchmarking**, because it involves identifying a peer group of similar organizations for each organization being evaluated. **Peer group benchmarking has face validity in the provider community and has been shown to reduce**

**the number of statistical outliers**, presumably because organizational characteristics (e.g., size, teaching status, ownership) explain some of the variation in outcomes that would otherwise be attributed to the individual organization.[146]

It is not always clear how a peer group of similar organizations should be constructed, as different approaches may yield different results.[147,148] At the extreme, some provider organizations might argue that they are unique in geography and structure, and therefore they do not have any peer group to which they can be compared. Even if an appropriate peer group can be identified, many question whether meaningful performance differences across different types of provider organizations should be "covered up" by attributing those differences to immutable organizational characteristics rather than to the organization itself.[149] Therefore, **most report card sponsors now benchmark provider organizations against all of the organizations with which they compete in a geographic market**, without regard to their size, volume, or teaching status.

### *Community Collaborative Example*

Both the **Wisconsin Health Care Value Exchange** and the **Michigan–Greater Detroit Area Health Council** are using AHRQ's NHQR to help determine their CVEs' performance improvement priorities. The report presents more than 220 measures at the State and national level and allows these CVEs to gauge their local challenges and opportunities. The Wisconsin CVE, which is a leader in data collection and measurement, used the NHQR to confirm many of its own findings. The Michigan CVE found that its State experience for asthma care was below average whereas its diabetes care was about average. Although asthma appeared to be the more problematic condition, the CVE chose to leverage its resources and address diabetes because the State of Michigan recently implemented an asthma care plan.

## Question 24. When and how should providers review data before public reports are released?

Studies show that provider acceptance of performance measurement and public reporting is largely dependent on the perceived validity of the measures.[142] **Most organizations pursuing performance measurement understand the necessity of measure validation and are working to integrate constructive provider feedback before reports are released**. This review and feedback may focus on one or more of three areas: (1) general analytic methods; (2) attribution of cases to specific providers; and (3) information about cases that would affect denominator exclusion, numerator determination, or risk adjustment.

Denominator exclusion relates to the concept of "exception reporting," which allows providers to identify specific patients who should not be eligible for inclusion in the quality measure. Numerator determination relates to whether the patient actually experienced an adverse outcome or actually failed to receive appropriate therapy. Risk adjustment relates to whether the data capture the patient's true severity of illness and therefore his or her true risk of an adverse outcome.

**Exception reporting** is a commonly used physician review method in the United Kingdom's pay-for-performance program, as well as The Joint Commission's Core Measures program. It was developed to allow providers to pursue quality improvement and avoid penalties for patients not meeting measure specifications for reasons that could not be captured in administrative data (e.g., newly diagnosed within the practice or had an allergy or other contraindication to treatment). Similarly, The Joint Commission "accepts" any physician statement in the record that a patient had a medical contraindication to the medication of interest, even if that contraindication is not supported by clinical evidence (www.qualitymeasures.ahrq.gov/summary/summary.aspx?ss=1&doc_id=773). Previous studies have explored concerns about providers "gaming" the system (shrinking the denominator by excluding patients who should be treated), with mixed conclusions.[150-152] Exception reporting may be particularly favored for MediCal beneficiaries due to provider perception of higher nonadherence and greater barriers to care in this population.

### *Mechanisms for Soliciting Provider Review and Feedback*

There are several possible mechanisms for soliciting provider review and feedback. Community quality collaboratives may choose to use one or more of these mechanisms, depending how much is already known about the quality of the data and the validity of the analyses based on those data. For example, when Medicaid claims data are used for reporting, continuous tracking and physician attribution may be problematic due to frequent eligibility changes within a 12-month period. **The mechanisms listed below are ordered from the most costly and time-consuming to the least.**

1. Send each provider (either routinely or upon request) a patient-level or claims-level data file summary with the specific cases and quality-related data elements attributed to him or her. **Allow a limited period (typically 21 days to 3 months) for providers to challenge the specific cases attributed to them, the quality-related data elements, or both**. A more limited version of this option, less susceptible to manipulation, would be to send only the attributed case list to each provider.
2. Send each provider (either routinely or upon request) a patient-level or claims-level data file with the specific cases and quality-related data elements attributed to him or her. **Allow a limited period (typically 21 days to 3 months) for providers to prepare a public response, without allowing them to challenge anything**.
3. Send each provider (either routinely or upon request) a draft copy of the report, Web materials, and other documents related to the planned public release. **Allow a limited period (typically 7-28 days) for providers to suggest specific changes to any of these documents**.
4. Send each provider (either routinely or upon request) a draft copy of the report, Web materials, and other documents related to the planned public release. **Allow a limited period (typically 7-28 days) for providers to review and prepare a public response**, which may accompany the final release.
5. Send each provider (either routinely or upon request) an advance copy of the report, Web materials, and other documents related to the planned public release. Do not solicit any suggestions or comments, but **alert providers that they should expect inquiries from media organizations and others**.

### *Community Collaborative Examples*

**Washington-Puget Sound Health Alliance** (WPSHA), The Alliance in Wisconsin, and the BQI Project provide examples of alternative provider review processes. WPSHA published the

*Reasonableness Review Process for Medical Groups*, which details how to access draft results for provider review and provide feedback through a secure online portal. They provided explanations of patient attribution and details about each measure as well as an appeals process. Specifically, “volunteer data suppliers” and medical groups worked together to confirm that specific measure results reflected a given clinic’s patients. Patients were reidentified for medical groups who then verified that the particular patient met the measure criteria and received a particular service from a particular clinician and clinic according to the measure specifications (www.wacommunitycheckup.org/editable/files/CommunityCheckup_Nov2008/ReasonablenessReviewProcessNov08.pdf).

Wisconsin’s purchaser coalition, **The Alliance**, sends results and documentation to hospitals before hosting a conference call during which technical specifications, including numerator/denominator definitions and risk-adjustment methods, are presented. Hospitals receive “rich” spreadsheets with far more detail than is publicly reported, at least 30 days before the scheduled release, and are encouraged to identify mistakes and raise concerns. On several occasions, hospitals have reported duplicate record submissions that altered their results. The Alliance corrected these mistakes and reported the proper results. Other issues that hospitals frequently raise relate to risk-adjustment methods and challenges to exclusions (e.g., hospital transfers for AHRQ Patient Safety Indicators or PSIs). The Alliance believes its methodology is sound and transparent and will only consider revisions if the method is technically invalid and has done so in the past after a hospital discovered bias in one of the PSI specifications.

**The six pilot sites in the Better Quality Information (BQI) project** invited physician feedback on patient attribution. In California, physicians were sent letters advising them to request data online and to contact the Pacific Business Group on Health (PBGH) to identify any errors. However, the Quality Improvement Organization confidentiality rules, which protect both patient and physician privacy, prohibited PBGH from sharing patient information that was not generated by the physician requesting the data. In this case, the auditing process did not work optimally for physicians with a high percentage of Medicare patients. The Massachusetts Chartered Value Exchange (CVE) offered an interactive Web-based tool so that providers could update their medical group affiliations. As part of the Quality Alliance Steering Committee’s High-Value Health Care project, a similar tool has been created so that physicians can review and correct the list of individual patients attributed to them. An active physician registry, when available, enhances proper patient attribution. In summary, the BQI pilot reaffirmed the importance of physician involvement to ensure proper attribution as well as to improve acceptance of the measurement process.

## Question 25. What are the critical success factors for selecting useful performance measures?

Established community quality collaboratives, such as Chartered Value Exchanges (CVEs), report several critical factors that contribute to a successful process of selecting measures and data.

1. It is important to have **healthy partnerships with diverse stakeholders** who support a common mission of performance measurement (i.e., pay for performance, public reporting, confidential reporting for quality improvement).

2. It is critical to **establish common goals** because those choices can affect the relative weights assigned to different evaluation criteria specified in Question 22. For example, a measure may be more relevant if a collaborative's primary goal is to inform consumer choice than if the primary purpose is to drive providers' quality improvement efforts.
3. **Continued, active engagement of key stakeholders** will help maintain support for the common mission established at the collaborative's inception, while allowing its goals to evolve over time as needed. Community quality collaboratives in the early stages of organizing should note that "key stakeholders" include consumer *and* provider representatives. Both offer viewpoints that are critical to the sustainability of the effort and the usability of the selected measures.

### *Successful Steps to Measure Selection*

Roski and Pawlson[153] suggest that after community quality collaboratives agree on the mission, they should:

1. **Address the goals or scope of measurement** (i.e., measure adherence to a single guideline or an assessment of the "quality" of care). More ambitious goals and a broader scope will inevitably necessitate more measures. An emphasis on accountability and transparency may lead to a larger set of measures, with more variable reliability, than an emphasis on improving consumer decisionmaking.
2. **Determine the number of measures required to meet the goal**, which varies according to desired level of validity and reliability and other technical performance characteristics. More measures are not necessarily better, if their reliability or validity is questionable, or if they are not relevant to the intended audience.
3. **Assess data source availability, reliability, and affordability** (e.g., electronic claims; pharmacy, laboratory, and medical records; paper records and surveys). Improving these features of the data may allow the goals or scope of measurement to expand, leading to a cycle of program improvement.

By following this three-step process, community quality collaboratives can ensure that they are making fully informed measure selections that are aligned with their specific goals and the needs of their stakeholders. In addition, they can ensure that their selections are consistent with the data and other resources available to them.

# Part V. Interpreting Quality and Resource Use Measures

## Question 26. How can quality and resource use measures be evaluated together to help identify high-value and low-value providers?

As community quality collaborative progress beyond their initial measure selection, data collection, and measurement, new challenges and opportunities will arise. One of those challenging opportunities is designing a useful construct for evaluating quality and resource use measures concurrently.

Very little research has been done in this area, although interest is building.[79] The Centers for Medicare & Medicaid Services (CMS) is currently exploring how to combine resource use metrics for episodes of care with quality metrics to differentiate physicians and to tie a portion of their payment to improvements or achievable benchmarks of efficiency. The CMS acute care episode (ACE) demonstration focuses on promoting efficiency by bundling all care delivered for an inpatient stay. Similar experiments are underway in the private sector; at least one health system is bundling payment for care in a hospital with the care delivered before and after the hospitalization for particular conditions.[114]

Figure 4 demonstrates one collaborative's approach, plotting a quality composite score with severity-adjusted hospital charges. In this example, providers falling in the upper-left quadrant score best in both quality and resource use. The Consumer-Purchaser Disclosure Project presented the diagram in Figure 5 to summarize the distribution of physicians in terms of "value" provided. Here, "value" is again constructed from both quality and resource use measures, but consumers and purchasers are encouraged to find physicians in the upper-right quadrant where high quality is delivered with low resource use.

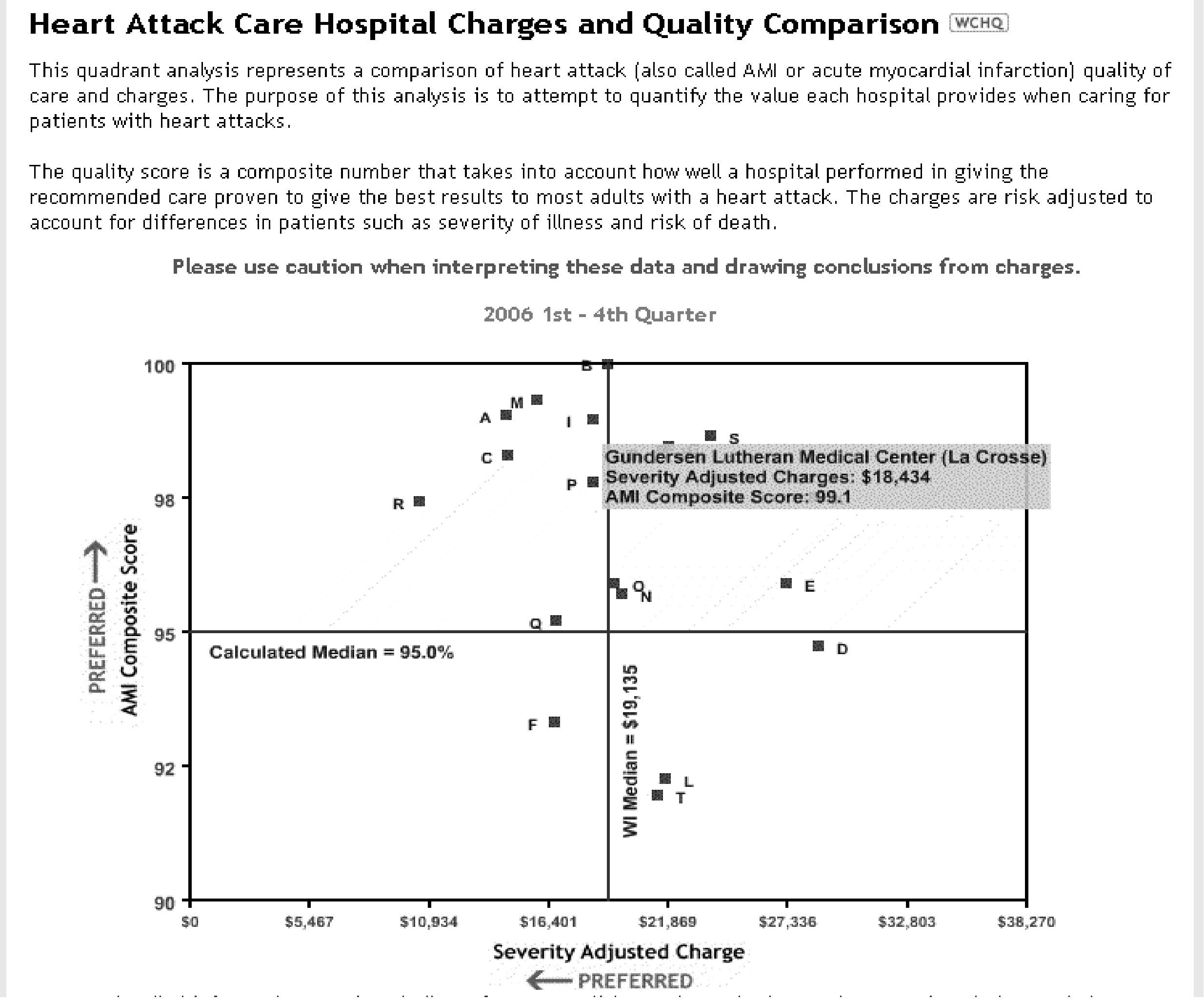
**Community quality collaboratives can serve as laboratories for both national and regional initiatives to package information on health care resource use and quality** in ways that will promote more efficient delivery of high-quality care. AHRQ has supported several useful tools to assist in this process, including:

1. The **"Talking Quality" Web site** at www.talkingquality.gov/ provides specific guidance about "what to say" and "how to say it" when communicating information on health care quality to consumers.
2. The **"Health Care Report Card Compendium"** at https://www.talkingquality.ahrq.gov/content/reportcard/search.aspx is a searchable directory of health care report cards that provide comparative information on the quality of health plans, hospitals, medical groups, individual physicians, and nursing homes.
3. **Evidence-based, empirically tested model reports** are available for public reporting of hospital performance on the AHRQ Quality Indicators site at: http://qualityindicators.ahrq.gov/downloads.htm#DraftModelReports. Similarly, templates for reporting on patient experience (based on the CAHPS® family of surveys) are available at: https://www.cahps.ahrq.gov/cahpskit/Reporting/reportingchoose.asp.
4. **MONAHRQ** (My Own Network, powered by AHRQ) is a Web-based application that will enable community collaboratives to input their own hospital administrative data and generate a data-driven Web site. MONAHRQ will be released in 2010.

5. “**Model Public Report Elements: A Sampler”** is an illustrative Web-based sampler of model public report elements. It spans five core Web pages that constitute a public report as well as functionality that facilitates use by consumers. The Sampler will be released in 2010 and accessible on AHRQ’s community quality collaboratives Web page at http://www.ahrq.gov/qual/value/localnetworks.htm.
6. **“Methodological Considerations in Generating Provider Performance Scores for Use in Public Reporting”** is a report focused on a set of 20 key methodological decisions associated with producing provider (e.g., hospital, physician, physician group) performance scores for use in public reporting. It also includes an explanation of the practical importance of each decision, a review of alternative decision paths, discussion of the pros and cons of each option, and examples from collaboratives. This resource will be released in 2010 and accessible on AHRQ’s community quality collaboratives Web page at http://www.ahrq.gov/qual/value/localnetworks.htm.

Similar tools and examples of successful dissemination approaches are likely to emerge from the Robert Wood Johnson Foundation’s **Aligning Forces for Quality** program (www.rwjf.org/qualityequality/af4q/focusareas/index.jsp). Through this program, 15 vanguard communities are bringing together key stakeholders to improve quality of care, measure and publicly report on quality of care, and engage consumers to make informed choices about their own health care. These efforts by both public and private funding agencies will provide a stronger evidence base for future initiatives to promote value-based purchasing and to increase consumer demand for high-value health care.

**Figure 4. Example of “value” plot diagram**



## Heart Attack Care Hospital Charges and Quality Comparison

This quadrant analysis represents a comparison of heart attack (also called AMI or acute myocardial infarction) quality of care and charges. The purpose of this analysis is to attempt to quantify the value each hospital provides when caring for patients with heart attacks.

The quality score is a composite number that takes into account how well a hospital performed in giving the recommended care proven to give the best results to most adults with a heart attack. The charges are risk adjusted to account for differences in patients such as severity of illness and risk of death.

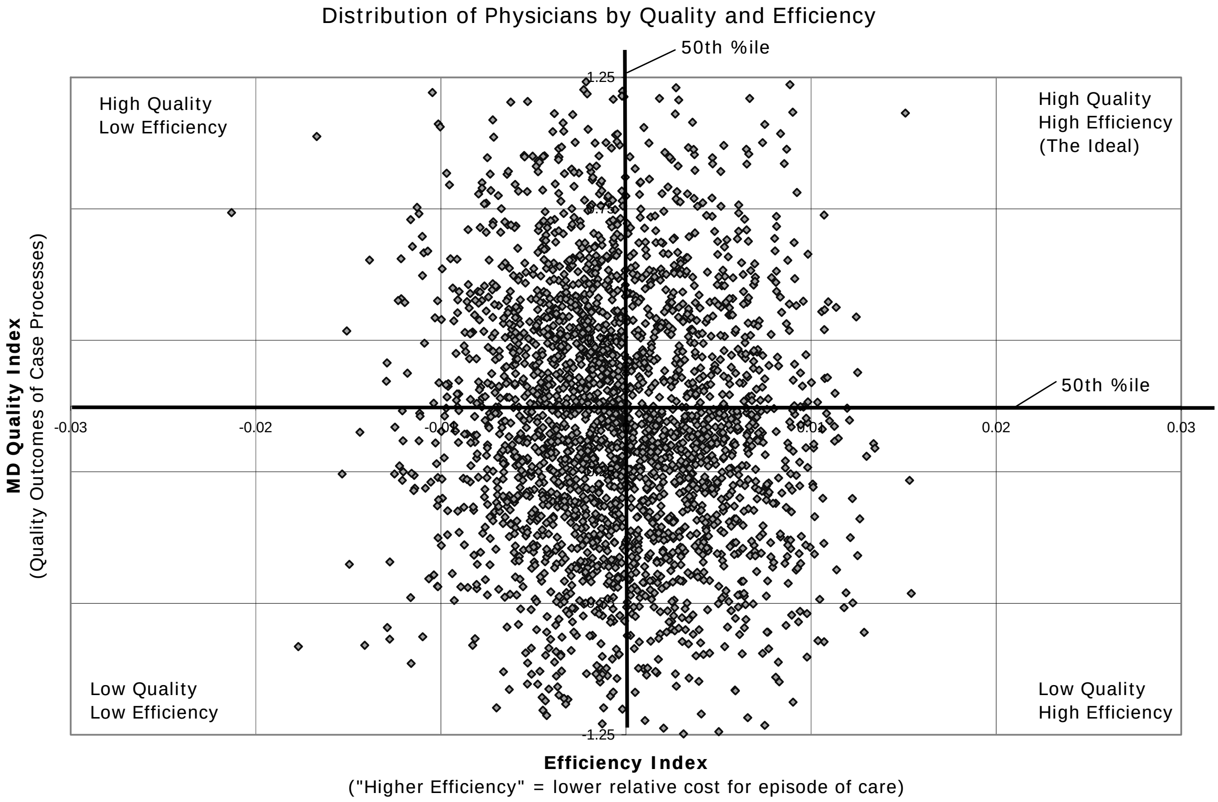
**Please use caution when interpreting these data and drawing conclusions from charges.**

2006 1st - 4th Quarter

Gundersen Lutheran Medical Center (La Crosse)
Severity Adjusted Charges: $18,434
AMI Composite Score: 99.1

**Source:** Wisconsin Collaboration for Healthcare Quality, 2009.

**Figure 5. Example of plotting physician value**



**Source:** Price & Cost Transparency: Understanding the Issues – Shaping the Agenda. Consumer-Purchaser Disclosure Project Invitational Working Session; May 25, 2006.

# References

1. Preparing for the 21st Century: focusing on quality in a changing health care system. Washington, DC: National Academy of Sciences; 1997. Available at: www.nas.edu/21st/health. Accessed November 30, 2009.
2. Chassin M, Galvin R. The urgent need to improve health care quality. Institute of Medicine National Roundtable on Health Care Quality. JAMA 1998 Sep 16;280(11):1000-5.
3. Institute of Medicine, Committee on Quality of Health Care in America. Crossing the quality chasm: a new health system for the 21st century. Washington, DC: National Academy Press; 2001.
4. Donaldson M. Measuring the quality of health care: a statement by the National Roundtable on Health Care Quality. Division of Health Care Services, Institute of Medicine. Washington, DC: National Academy Press; 1999.
5. Davis K. Room for improvement: patients report on the quality of their health care. Washington, DC: Commonwealth Fund, Program on Health Care Quality Improvement; 2002.
6. Health care quality in California: a primer; The case for quality in health care; Variations in medical practice; Underuse and overuse of medical services; Medical errors; Suggested readings on health care quality. Oakland: California HealthCare Foundation. Available at: www.chcf.org/publications/. Accessed May 1, 2009.
7. Snapshot: Health Care Costs 101. Oakland: California HealthCare Foundation; June 2005. Available at: www.allhealth.org/BriefingMaterials/californiahealthcarecosts2005-129.pdf. Accessed May 3, 2009.
8. Von Korff M, Barlow W, Cherkin D, et al. Effects of practice style in managing back pain. Ann Intern Med 1994 Aug 1;121(3):187-95.
9. Jiang HJ, Friedman B, Begun JW. Factors associated with high-quality/low-cost hospital performance. J Health Care Finance 2006 Spring;32(3):39-52.
10. Health care at the crossroads: development of a national performance measurement data strategy. Oakbrook Terrace, IL: The Joint Commission; 2008. Available at: www.jointcommission.org/NR/rdonlyres/24812778-7DBB-4F07-884C-9821FFD7792E/0/National_Perf_Measure_Data.pdf.
11. The Foundation of Research and Education of the American Health Information Management Association and the Medical Group Management Association Center for Research. AHRQ Conference on Health Care Data Collection and Reporting: Collecting and Reporting Data for Performance Measurement: Moving Toward Alignment. Rockville, MD: AHRQ; 2007. AHRQ Publication No. 07-0033-EF. Available at: http://healthit.ahrq.gov/portal/server.pt/gateway/PTARGS_0_1248_227079_0_0_18/AHRQ_DataReport_final.pdf. Accessed April 15 2009.
12. Kazandjian V. The epidemiology of quality. 1st ed. Gaithersburg, MD: Aspen Publishers; 1995.
13. Iezzoni LI. Assessing quality using administrative data. Ann Intern Med. 1997 Oct 15;127(8 Pt 2):666-74.
14. Joint Commission on Accreditation of Healthcare Organizations. Strategies to ease data collection burden. Hosp Peer Rev 2006 Jan;31(1):6, 11.
15. Lorence D, Ibrahim I. Benchmarking variation in coding accuracy across the United States. J Health Care Finance 2003;29(4):29.
16. Hsia D, Ahern C, Ritchie B, et al. Medicare reimbursement accuracy under the prospective payment system, 1985 to 1988. JAMA 1992;268(7):896.
17. Fisher E, Whaley F, Krushat W, et al. The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. Am J Pub Hlth 1992;82(2):243.
18. Steinwachs DM, Stuart ME, Scholle S, et al. A comparison of ambulatory Medicaid claims to medical records: a reliability assessment. Am J Med Qual 1998 Summer;13(2):63-69.
19. Pawlson LG, Scholle SH, Powers A. Comparison of administrative-only versus administrative plus chart review data for reporting HEDIS hybrid measures. Am J Manage Care 2007 Oct;13(10):553-8.
20. HEDIS 2006 CAHPS 3.0H survey file layouts. Washington, DC: National Committee on Quality Assurance; 2005. Available at: http://www.ncqa.org/tabid/439/Default.aspx.
21. Corriol C, Daucourt V, Grenier C, et al. How to limit the burden of data collection for quality indicators based on medical records? The COMPAQH experience. BMC Health Serv Res 2008;8:215.
22. Kerr EA, Hofer TP, Hayward RA, et al. Quality by any other name?: a comparison of three profiling systems for assessing health care quality. Health Serv Res 2007 Oct;42(5):2070-87.
23. Kaplan SHN, Sharon-Lise T. . Conceptual and analytical issues in creating composite measures of ambulatory care performance. Washington, DC: Robert Wood Johnson Foundation; 2006. RWJF Grant Number 51553.

24. Healthcare Cost and Utilization Project. Enhancing the clinical content of administrative data. Rockville, MD: AHRQ; September 2008. Available at: www.hcup-us.ahrq.gov/reports/clinicaldata.jsp. Accessed February 17, 2009.
25. Fry DE PM, Jordan HS, Hoaglin DC, et al. The hazards of using administrative data to measure surgical quality. Am Surgeon 2006;72(11):1031-7.
26. Pine M, Jordan HS, Elixhauser A, et al. Enhancement of claims data to improve risk adjustment of hospital mortality. JAMA 2007 Jan 3;297(1):71-76.
27. Fry DE PM, Jordan HS, Elixhauser A, et al. Combining administrative and clinical data to stratify surgical risk. Ann Surg 2007;246(5):875-85.
28. Yu F, Menachemi N, Berner E, et al. Full implementation of computerized physician order entry and medication-related quality outcomes: a study of 3364 hospitals. Am J Med Qual 2009;24(4):278.
29. Jha A, DesRoches C, Campbell E, et al. Use of electronic health records in U.S. hospitals. N Engl J Med 2009;360(16):1628.
30. Bartschat W, Burrington-Brown J, et al. Surveying the RHIO landscape. A description of current RHIO models, with a focus on patient identification. J AHIMA 2006 Jan;77(1):64A-D.
31. Jha AK, Doolan D, Grandt D, et al. The use of health information technology in seven nations. Int J Med Inform 2008 Dec;77(12):848-54.
32. DesRoches C, Campbell E, Rao S, et al. Electronic health records in ambulatory care—a national survey of physicians. N Engl J Med 2008;359(1):50.
33. Greene J. The trials and tribulations of health information sharing: the turbulent rise of the RHIO. Ann Emerg Med 2007 Nov;50(5):549-51.
34. Nationwide Health Information Network (NHIN): history and background. Washington, DC: U.S. Department of Health and Human Services. Available at: http://healthit.hhs.gov/portal/server.pt?open=512&mode=2&cached=true&objID=1142. Accessed April 15, 2009.
35. Indiana Health Information Exchange Quality Health First. Available at: www.ihie.com/. Accessed January 23, 2009.
36. White paper: why California is ready for statewide health information exchange. San Francisco: California Regional Health Infomation Organization. Available at: www.calrhio.org/crweb-files/docs-hie/White%20Paper%20-%20Statewide%20California%20HIE.pdf. Accessed April 15, 2009.
37. Harrison JH, Jr., Aller RD. Regional and national health care data repositories. Clin Lab Med 2008 Mar;28(1):101-17, vii.
38. Berndt DF, Fisher JW, Hevner AR, et al. Healthcare data warehousing and quality assurance. Computer 2001 Dec;34(12):56-65.
39. Lazarus R, Yih K, Platt R. Distributed data processing for public health surveillance. BMC Pub Hlth 2006 Sep 19;6:235-46.
40. Monitoring and evaluating Medicaid fee-for-service care management programs. User's guide. Rockville, MD: AHRQ; November 2007. AHRQ Publication No. 08-0012. Available at: www.ahrq.gov/qual/medicaidffs/. Accessed April 15, 2009.
41. National Committee on Quality Assurance. HEDIS Compliance Audit Program. Available at: www.ncqa.org/tabid/205/Default.aspx. Accessed April 15, 2009.
42. The Joint Commission. Accreditation programs: hospitals. Available at: www.jointcommission.org/AccreditationPrograms/Hospitals/. Accessed April 15, 2009.
43. Quality Net. Data Validation Overview: Hospitals. Available at: www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1140537255912. Accessed April 15, 2009.
44. Enhancing physician quality performance measurement and reporting through data aggregation: the Better Quality Information (BQI) to Improve Care for Medicare Beneficiaries Project, Final Report. Baltimore, MD: Centers for Medicare & Medicaid Services. Available at: www.cms.hhs.gov/bqi. Accessed February 6, 2009.
45. Integrated Healthcare Association. Pay for Performance Program audit review guidelines: measurement year 2009. Washington, DC: NCQA; November 2009. Available at: www.iha.org/pdfs_documents/p4p_california/MY%202009%20P4P%20Audit%20Manual_pp%201-50_final.pdf.
46. Centers for Medicare & Medicaid Services. HIPAA – General Information. 2005. Available at: www.cms.hhs.gov/HIPAAGenInfo/. Accessed April 15, 2009.

47. U.S. Department of Health and Human Services. Summary of the HIPAA Privacy Rule. Last revised May 2003. Available at: www.hhs.gov/ocr/privacy/hipaa/understanding/summary/privacysummary.pdf.
48. The Federal Privacy Act of 1974 and HIPAA Privacy Rule of 1996: a comparison. Sacramento: California Department of Veterans Affairs; September 2007.
49. Candace G. Understanding and complying with HIPAA. J Perianesth Nurs 2003;18(3):182-5.
50. Quality improvement organizations and health information exchange. Washington, DC: American Health Quality Foundation; March 6, 2006; Available at: www.ahqa.org/pub/uploads/qio_hie_final_report_march_6_2006.pdf Accessed April 17, 2009.
51. New OCR guidance on the HIPAA privacy rule and the electronic exchange of health information. Baltimore, MD: Centers for Medicare & Medicaid Services. Available at: www.hhs.gov/ocr/hipaa/hit/. Accessed March 03, 2009.
52. Rosenbaum S, Kornblet S, Borzi PC. An assessment of legal issues raised in “high performing” health plan quality and efficiency tiering arrangements: can the patient be saved? Health Pol Rep 2007 Oct 8.
53. Inpatient Quality Indicators: technical specifications. Rockville, MD: AHRQ; June 2002. Accessed April 17, 2009.
54. Houchens RL, Elixhauser A, Romano PS. How often are potential patient safety events present on admission? Jt Comm J Qual Patient Saf 2008 Mar;34(3):154-63.
55. CAHPS: Surveys and Tools to Advance Patient-Centered Care. Rockville, MD: AHRQ. Available at: www.cahps.ahrq.gov/default.asp. Accessed April 16, 2009.
56. Prevention Quality Indicators Overview. Rockville, MD: AHRQ; July 2004. Available at: www.qualityindicators.ahrq.gov/pqi_overview.htm. Accessed April 16, 2009.
57. Jewett J, Hibbard J. Comprehension of quality care indicators: differences among privately insured, publicly insured, and uninsured. Health Care Financ Rev 1996 Fall;18(1):75-94.
58. Khuri SF, Daley J, Henderson W, et al. The Department of Veterans Affairs NSQIP: the first national, validated, outcome-based, risk-adjusted, and peer-controlled program for the measurement and enhancement of the quality of surgical care. National VA Surgical Quality Improvement Program. Ann Surg 1998 Oct;228(4):491-507.
59. Neumayer L, Mastin M, Vanderhoof L, et al. Using the Veterans Administration National Surgical Quality Improvement Program to improve patient outcomes. J Surg Res 2000 Jan;88(1):58-61.
60. Greenfield S, Kaplan SH, Kahn R, et al. Profiling care provided by different groups of physicians: effects of patient case-mix (bias) and physician-level clustering on quality assessment results. Ann Intern Med 2002 Jan 15;136(2):111-21.
61. Scholle SH, Roski J, Adams JL, et al. Benchmarking physician performance: reliability of individual and composite measures. Am J Manag Care 2008 Dec;14(12):833-8.
62. Hofer TP, Hayward RA, Greenfield S, et al. The unreliability of individual physician “report cards” for assessing the costs and quality of care of a chronic disease. JAMA 1999 Jun 9;281(22):2098-2105.
63. Kaplan SH, Griffith JL, Price LL, et al. Improving the reliability of physician performance assessment: identifying the “physician effect” on quality and creating composite measures. Med Care 2009 Apr;47(4):378-87.
64. Luft HS, Romano PS. Chance, continuity, and change in hospital mortality rates. Coronary artery bypass graft patients in California hospitals, 1983 to 1989. JAMA 1993 Jul 21;270(3):331-7.
65. Dimick JB, Welch HG, Birkmeyer JD. Surgical mortality as an indicator of hospital quality: the problem with small sample size. JAMA 2004 Aug 18;292(7):847-51.
66. Physician Consortium for Performance Improvement. Chicago, IL: American Medical Association. Available at: www.ama-assn.org/ama/pub/physician-resources/clinical-practice-improvement/clinical-quality/physician-consortium-performance-improvement.shtml. Accessed April, 19, 2009.
67. Talking Quality. Glossary and acronyms. Rockville, MD: AHRQ. Available at: https://www.talkingquality.ahrq.gov/content/reportcard/glossary.aspx. Accessed April 27, 2009.
68. Shaller D, Sofaer S, Findlay SD, et al. Consumers and quality-driven health care: a call to action. Health Aff (Millwood) 2003 Mar-Apr;22(2):95-101.
69. Shwartz M, Ash AS. Composite measures: matching the method to the purpose. Rockville, MD: AHRQ. Available at: www.qualitymeasures.ahrq.gov/resources/commentary.aspx?file=Composite.inc. Accessed April 16, 2009.
70. Jha AK, Li Z, Orav EJ, et al. Care in U.S. hospitals--the Hospital Quality Alliance Program. N Engl J Med 2005 Jul 21;353(3):265-74.

71. Gandhi TK, Francis EC, Puopolo AL, et al. Inconsistent report cards: assessing the comparability of various measures of the quality of ambulatory care. Med Care 2002 Feb;40(2):155-65.
72. Feinstein AR. Multi-item "instruments" vs. Virginia Apgar's principles of clinimetrics. Arch Intern Med 1999 Jan 25;159(2):125-8.
73. Inpatient Quality Indicators (IQI) Composite Measure Workgroup final report. Rockville, MD: AHRQ; 2008. Available at: www.qualityindicators.ahrq.gov/downloads/iqi/AHRQ_IQI_Workgroup_Final.pdf
74. Caldis T. Composite health plan quality scales. Health Care Financ Rev 2007 Spring;28(3):95-107.
75. Composite measure evaluation framework and national voluntary consensus standards for mortality and safety: composite measures. Washington, DC: National Quality Forum; August 2009; Available at: www.qualityforum.org/Projects/c-d/Composite_Evaluation_Framework/Composite_Evaluation_Framework_and_Composite_Measures.aspx.
76. Zaslavsky AM, Shaul JA, Zaborski LB, et al. Combining health plan performance indicators into simpler composite measures. Health Care Financ Rev 2002 Summer;23(4):101-15.
77. Reporting Measures for the CAHPS Clinician & Group Survey. CAHPS Clinician & Group Survey and Reporting Kit. Rockville, MD: AHRQ; April 2007. Available at: www.cahps.ahrq.gov/CAHPSkit/files/309_CG_Reporting_Measures.htm. Accessed April 27, 2009.
78. Solomon LS, Hays RD, Zaslavsky AM, et al. Psychometric properties of a group-level Consumer Assessment of Health Plans Study (CAHPS) instrument. Med Care 2005 Jan;43(1):53-60.
79. Hargraves JL, Hays RD, Cleary PD. Psychometric properties of the Consumer Assessment of Health Plans Study (CAHPS) 2.0 adult core survey. Health Serv Res 2003 Dec;38(6 Pt 1):1509-27.
80. Reeves D, Campbell SM, Adams J, et al. Combining multiple indicators of clinical quality: an evaluation of different analytic approaches. Med Care 2007 Jun;45(6):489-96.
81. Nolan T, Berwick DM. All-or-none measurement raises the bar on performance. JAMA 2006 Mar 8;295(10):1168-70.
82. Timbie JW, Shahian DM, Newhouse JP, et al. Composite measures for hospital quality using quality-adjusted life years. Stat Med 2009 Apr 15;28(8):1238-54.
83. Zhan C, Miller MR. Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization. JAMA 2003 Oct 8;290(14):1868-74.
84. Rivard PE, Luther SL, Christiansen CL, et al. Using patient safety indicators to estimate the impact of potential adverse events on outcomes. Med Care Res Rev 2008 Feb;65(1):67-87.
85. Glickman S, Boulding W, Roos J, et al. Alternative pay-for-performance scoring methods: implications for quality improvement and patient outcomes. Med Care 2009;47(10):1062.
86. Premier Hospital Quality Incentive Demonstration. Baltimore, MD: Centers for Medicare & Medicaid Services.Available at: http://www.cms.hhs.gov/HospitalQualityInits/35_hospitalpremier.asp. Accessed April 16, 2009.
87. Scott IA, Coory MD, Harper CM. The effects of quality improvement interventions on inhospital mortality after acute myocardial infarction. Med J Aust 2001 Nov 5;175(9):465-70.
88. Scott I, Youlden D, Coory M. Are diagnosis specific outcome indicators based on administrative data useful in assessing quality of hospital care? Qual Saf Health Care 2004 Feb;13(1):32-39.
89. Krumholz HM, Wang Y, Mattera JA, et al. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. Circulation 2006 Apr 4;113(13):1683-92.
90. Shahian DM, Normand SL. Comparison of "risk-adjusted" hospital outcomes. Circulation 2008 Apr 15;117(15):1955-63.
91. Quality Net. Methodologies. Available at: https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1163010421830. Accessed April 16, 2009.
92. Stukenborg GJ, Wagner DP, Harrell FE, Jr., et al. Which hospitals have significantly better or worse than expected mortality rates for acute myocardial infarction patients? Improved risk adjustment with present-at-admission diagnoses. Circulation 2007 Dec 18;116(25):2960-8.
93. Pine M, Jordan HS, Elixhauser A, et al. Modifying ICD-9-CM coding of secondary diagnoses to improve risk-adjustment of inpatient mortality rates. Med Decis Making 2009 Jan-Feb;29(1):69-81.
94. Iezzoni L. Risk adjustment for measuring health care outcomes. 3rd ed. Chicago: Health Administration Press; 2003.
95. O'Malley AJ, Zaslavsky AM, Elliott MN, et al. Case-mix adjustment of the CAHPS Hospital Survey. Health Serv Res 2005 Dec;40(6 Pt 2):2162-81.

96. CAHPS Hospital Survey, Mode & Patient Mix Adjustment. Available at: www.hcahps.org/modeadjustment.aspx#mix. Accessed March 14, 2009.
97. Zaslavsky A, Zaborski L, Ding L, et al. Adjusting performance measures to ensure equitable plan comparisons. Health Care Financ Rev 2001;22(3):109-26.
98. Kim M, Zaslavsky AM, Cleary PD. Adjusting pediatric Consumer Assessment of Health Plans Study (CAHPS) scores to ensure fair comparison of health plan performances. Med Care 2005 Jan;43(1):44-52.
99. Zaslavsky AM, Cleary PD. Dimensions of plan performance for sick and healthy members on the Consumer Assessments of Health Plans Study 2.0 survey. Med Care 2002 Oct;40(10):951-64.
100. Zaslavsky AM, Landon BE, Beaulieu ND, et al. How consumer assessments of managed care vary within and among markets. Inquiry 2000 Summer;37(2):146-61.
101. Romano PS. Should health plan quality measures be adjusted for case mix? Med Care 2000 Oct;38(10):977-80.
102. Fiscella K, Franks P, Gold MR, et al. Inequality in quality: addressing socioeconomic, racial, and ethnic disparities in health care. JAMA 2000 May 17;283(19):2579-84.
103. Agency for Healthcare Research and Quality, American Institutes for Research, Harvard Medical School, RAND Corporation. The CAHPS® Clinician & Group Survey. Washington, DC: NQF; July 13, 2006; Available at: www.aqaalliance.org/October24Meeting/PerformanceMeasurement/CAHPSCGtext.doc.
104. Quality insights 2007: patient experiences in primary care, massachusetts statewide results. Watertown: Massachusetts Health Quality Partners. Available at: www.mhqp.org/quality/pes/pesMASumm.asp?nav=031600. Accessed April 17, 2009.
105. Davis C. Personal Interview: Missouri/Kansas-Kansas City Quality Improvement Consortium. Kansas City, Missouri, March 12, 2009.
106. Adams K, Burstin H. Measurement framework: evaluating efficiency across patient-focused episodes of care. Washington, DC: NQF; January 12, 2009. Available at: http://qualityforum.org/Projects/Episodes_of_Care_Framework.aspx. Accessed April 17, 2009.
107. McGlynn EA. Final report: identifying, categorizing, and evaluating health care efficiency measures. Rockville, MD: AHRQ; April 2008. AHRQ Publication No. 08-0030. Available at: www.ahrq.gov/qual/efficiency/index.html. Accessed April 17, 2009.
108. Valdmanis VG, Rosko MD, Mutter RL. Hospital quality, efficiency, and input slack differentials. Health Serv Res 2008 Oct;43(5 Pt 2):1830-48.
109. Healthcare Cost & Utilization Project. Rockville, MD: AHRQ. Available at: www.ahrq.gov/data/hcup/. Accessed April 18, 2009.
110. Lave JR, Pashos CL, Anderson GF, et al. Costing medical care: using Medicare administrative data. Med Care 1994 Jul;32(7 Suppl):JS77-89.
111. Friedman B, De La Mare J, Andrews R, et al. Estimating hospital cost for discharged patients: practical options in an era of data restrictions. J Health Care Finance 2002;29:1-13.
112. Milliman RBRVS for Hospitals. Milliman White Papers. Seattle: Milliman; February 2009. Available at: www.milliman.com/expertise/healthcare/products-tools/rbrvs/pdfs/milliman-rbrvs-for-hospitals-WP02-01-09.pdf. Accessed April 27, 2009.
113. Milstein A, Lee TH. Comparing physicians on efficiency. N Engl J Med 2007 Dec 27;357(26):2649-52.
114. Medicare resource use measurement plan. Baltimore, MD: Centers for Medicare & Medicaid Services. Available at: www.cms.hhs.gov/QualityInitiativesGenInfo/downloads/ResourceUse_Roadmap_OEA_1-15_508.pdf. Accessed April 17, 2009.
115. 2008 National Healthcare Quality & Disparities Reports. Rockville, MD: Agency for Healthcare Research and Quality. Available at: www.ahrq.gov/qual/qrdr08.htm. Accessed November 23, 2009.
116. Mutter RL, Rosko MD, Wong HS. Measuring hospital inefficiency: the effects of controlling for quality and patient burden of illness. Health Serv Res 2008 Dec;43(6):1992-2013.
117. Quality Alliance Steering Committee. Available at: www.healthqualityalliance.org/. Accessed April 17, 2009.
118. “Patient charter” for physician performance measurement, reporting and tiering programs. San Francisco: Consumer Purchaser Disclosure Project.Available at: http://healthcaredisclosure.org/activities/charter/. Accessed April 17, 2009.
119. Lake T, Colby M, Peterson S. Health plans’ use of physician resource use and quality measures final report. Washington, DC: Medicare Payment Advisory Commission; October 24, 2007; Available at: www.medpac.gov/documents/6355%20MedPAC%20Final%20Report%20with%20Appendices%201-24-08.pdf.

120. Report to the Congress: increasing the value of Medicare. Washington, DC: Medicare Payment Advisory Commission; 2006. Available at: www.medpac.gov/documents/Jun06_EntireReport.pdf. Accessed April 17, 2009.
121. California HealthCare Foundation, Manatt Health Solutions, Robert Mittman. An unprecedented opportunity: using Federal stimulus funds to advance health IT in California. Oakland: California HealthCare Foundation; 2009. Available at: www.chcf.org/topics/view.cfm?itemid=133864. Accessed April 17, 2009.
122. National Priorities Partnership. National priorities and goals: aligning our efforts to transform America's healthcare. Washington, DC: National Quality Forum;2008. Available at: www.nationalprioritiespartnership.org/uploadedFiles/NPP/08-253-NQF%20ReportLo%5B6%5D.pdf. Accessed April 17, 2009.
123. Painter MW, Lavizzo-Mourey R. Aligning Forces for Quality: a program to improve health and health care in communities across the United States. Health Aff (Millwood) 2008 Sep-Oct;27(5):1461-3.
124. Flowchart on the Quality Measurement Enterprise. Washington, DC: Hospital Quality Alliance. Available at: www.hospitalqualityalliance.org/hospitalqualityalliance/resources/resources_page.html. Accessed April 17, 2009.
125. Donabedian A. Explorations in quality assessment and monitoring. Vol. 3, The methods and findings of quality assessment and monitoring: an illustrated analysis. Chicago: Health Administration Press; 1985.
126. Donabedian A. Explorations in quality assessment and monitoring. Chicago: Health Administration Press; 1982.
127. Donabedian A. The definition of quality and approaches to its assessment. Chicago: Health Administration Press; 1980.
128. Quality Measures Development Overview. Baltimore, MD: Centers for Medicare & Medicaid Services.Available at: www.cms.hhs.gov/QualityInitiativesGenInfo/downloads/QualityMeasuresDevelopmentOverview.pdf. Accessed April 17, 2009.
129. HEDIS Life Cycle. Washington, DC: NCQA. Available at: www.ncqa.org/tabid/425/Default.aspx. Accessed April 17 2009.
130. Measure Evaluation Criteria. Washington, DC: NQF; August 2008. Available at: www.qualityforum.org/docs/measure_evaluation_criteria.aspx. Accessed April 17, 2009.
131. Consensus Development Process. Washington, DC: NQF. Available at: www.qualityforum.org/Measuring_Performance/Consensus_Development_Process.aspx. Accessed April 17, 2009.
132. Medicare Program; Changes to the hospital inpatient prospective payment systems and fiscal year 2009 rates; payments for graduate medical education in certain emergency situations; changes to disclosure of physician ownership in hospitals and physician self-referral rules; updates to the long-term care prospective payment system; updates to certain IPPS-excluded hospitals; and collection of information regarding financial relationships between hospitals; final rules. 73 Fed Reg 48433-49083; August 19, 2008. Available at: www.gpo.gov/fdsys/pkg/FR-2008-08-19/html/E8-17914.htm.
133. AQA parameters for selecting measures for physician and other clinician performance. Washington, DC: AQA Alliance. Available at: www.aqaalliance.org/performancewg.htm. Accessed April 17, 2009.
134. Physician Consortium for Performance Improvement. Guidance for requesting Physician Consortium for Performance Improvement® (PCPI) review and approval of measures developed independently by PCPI voting members. Approved October 5, 2007. Available at: www.ama-assn.org/ama1/pub/upload/mm/370/meas_dev_ind_pcpi.pdf. Accessed April 17, 2009.
135. Roski J, Kim MG. Current efforts of regional and national performance measurement initiatives around the United States. Am J Med Qual 2009 Sep 3 (Epub ahead of print).
136. Guidance for using the AHRQ quality indicators for public reporting or payment, Appendix B: Public reporting evaluation framework—comparison of recommended evaluation criteria in five existing national frameworks. Rockville, MD: Agency for Healthcare Research and Quality; 2009. Available at: http://qualityindicators.ahrq.gov/downloads/technical/qi_guidance_appendix_B.pdf. Accessed April 19, 2009.
137. National Healthcare Quality Report. Rockville, MD: Agency for Healthcare Research and Quality; 2008. AHRQ Publication No. 08-0040. Available at: www.ahrq.gov/qual/nhqr07/nhqr07.pdf. Accessed April 17, 2009.
138. Remus D, Fraser I. Guidance for using the AHRQ quality indicators for hospital-level public reporting or payment. Rockville, MD: Agency for Healthcare Research and Quality; 2004. AHRQ Publication No. 04-

0086-EF. Available at: www.qualityindicators.ahrq.gov/archives/documents/qi_guidance.pdf. Accessed April 17, 2009.

139. National Quality Measures Clearinghouse, Resources - Measure Validity. Rockville, MD: Agency for Healthcare Research and Quality. Available at: http://qualitymeasures.ahrq.gov/resources/measure_domains.aspx Accessed April 19, 2009.
140. Bott J. Personal interview: Value Based Purchasing Manager, The Alliance. Ritley D, Madison, Wisconsin, March 12, 2009.
141. Weissman NW, Allison JJ, Kiefe CI, et al. Achievable benchmarks of care: the ABCs of benchmarking. J Eval Clin Pract 1999 Aug;5(3):269-81.
142. Kiefe CI, Weissman NW, Allison JJ, et al. Identifying achievable benchmarks of care: concepts and methodology. Int J Qual Health Care 1998 Oct;10(5):443-7.
143. Wessell AM, Liszka HA, Nietert PJ, et al. Achievable benchmarks of care for primary care quality indicators in a practice-based research network. Am J Med Qual 2008 Jan-Feb;23(1):39-46.
144. Wennberg JE. Understanding geographic variations in health care delivery. N Engl J Med 1999 Jan 7;340(1):52-53.
145. Dudley RA, Rosenthal MB. Pay for performance: a decision guide for purchasers. Rockville, MD: AHRQ; 2006. AHRQ Publication No. 06-0047. Available at: www.ahrq.gov/qual/p4pguide.pdf. Accessed April 19, 2009.
146. Austin PC, Alter DA, Anderson GM, et al. Impact of the choice of benchmark on the conclusions of hospital report cards. Am Heart J 2004 Dec;148(6):1041-6.
147. Byrne MM, Daw CN, Nelson HA, et al. Method to develop health care peer groups for quality and financial comparisons across hospitals. Health Serv Res 2009 Apr;44(2 Pt 1):577-92.
148. Zodet MW, Clark JD. Creation of hospital peer groups. Clin Perform Qual Health Care 1996 Jan-Mar;4(1):51-57.
149. Romano PS. Peer group benchmarks are not appropriate for health care quality report cards. Am Heart J 2004 Dec;148(6):921-3.
150. Fleetcroft R, Steel N, Cookson R, et al. “Mind the gap!” Evaluation of the performance gap attributable to exception reporting and target thresholds in the new GMS contract: National database analysis. BMC Health Serv Res 2008;8:131.
151. Doran T, Fullwood C, Reeves D, et al. Exclusion of patients from pay-for-performance targets by English physicians. N Engl J Med 2008 Jul 17;359(3):274-84.
152. Conway PH, Clancy C. Transformation of health care at the front line. JAMA 2009 Feb 18;301(7):763-5.
153. Roski JP, Pawlson G. Putting reliable health care performance measurement systems into practice. NCQA Issue Brief. Washington, DC: NCQA; 2007. Available at: www.ncqa.org/Portals/0/Publications/Resource%20Library/Issue%20Briefs/Putting_Peformance_Measures_into_Practice.pdf.

# Appendix A

## Measure Evaluation Framework

Appendix A presents a matrix of measure evaluation criteria* used by five major national organizations. The matrix is borrowed from AHRQ's *Guidance for Using the AHRQ Quality Indicators for Public Reporting or Payment* - *Appendix B: Public Reporting Evaluation Framework*. Please note that terminology may vary by framework.

| Public Reporting Evaluation Framework | | | | | |
|---|---|---|---|---|---|
| **Evaluation criteria and requirements** | **National Quality Forum** | **AHRQ Quality Indicators** | **National Healthcare Quality Report** | **The Joint Commission** | **National Committee for Quality Assurance** |
| ***1. Importance*** | **Important** | **—Face validity**<br>**—Foster real quality improvement** | **Importance** | | **Relevance** |
| • Assesses an important leverage point for improving quality; significant to target audiences; impact on health<br>• Opportunity for improvement, considerable variation in quality of care exists<br>• Aspect of quality is under provider or health system control<br>• Should not create incentives or rewards to improve without truly improving quality of care | • Leverage point for improving quality<br>• Considerable variation in quality of care exists<br>• Performance in the area is suboptimal<br>• Aspect of quality is under provider or health system control[1] | • Measure an important aspect of quality that is subject to provider or health system control<br>• Should not create incentives or rewards to improve without truly improving quality of care | • Impact on health<br>• Meaningfulness<br>• Susceptibility to being influenced by health care | • Targets improvement in the health of populations<br>• Under provider control | • Strategic importance<br>• Health importance<br>• Meaningfulness to decision makers<br>• Variance among systems<br>• Potential for improvement<br>• Controllability<br>• Financial importance |

*Criteria are taken verbatim from the various sources and have not been edited.

---

[1] This criterion is in the NQF framework at the scope/priority level and not at the individual measure evaluation level.

| Evaluation criteria and requirements | National Quality Forum | AHRQ Quality Indicators | National Healthcare Quality Report | The Joint Commission | National Committee for Quality Assurance |
|---|---|---|---|---|---|
| ***2. Scientific acceptability*** | **Scientifically acceptable** | **—Precision**<br>**—Minimum bias**<br>**—Construct validity** | **Scientific soundness** | | **Scientific soundness** |
| • Relationship to quality is based on scientific evidence<br>• Well defined and precisely specified<br>• Valid, measures the intended aspect of quality; accurately represents the concept being evaluated; data sources are comparable<br>• Adequate proportion of total variation is explained by provider performance and amount of variation in measurement is small after provider performance and patient characteristics are taken into account<br>• Reliable, producing the same results a high proportion of time in the same population<br>• Precise, adequately discriminating between real differences in provider performance and reasonable sample size exists to detect actual differences; captures all possible cases and bias related to case exclusion or limited data are minimal.<br>• Risk adjustment is adequate to address confounding bias | • Well defined and precisely specified<br>• Reliable, producing the same results a high proportion of time in the same population<br>• Valid, accurately representing the concept being evaluated<br>• Precise, adequately discriminating between real differences in provider performance<br>• Adaptable to patient preferences and variety of settings<br>• Adequate and specified risk adjustment strategy exists<br>• Evidence is available linking process measures to outcomes | • Have relatively large variation among providers that is not due to random variation or patient characteristics<br>• Should not be affected by systematic differences in patient case-mix<br>• When systematic differences exist, an adequate risk adjustment system is available based on HCUP discharge data<br>• Supported by evidence of a relationship to quality<br>• Related to other indicators intended to measure the same or related aspects of quality | • Explicitness of the evidence base<br>• Reliability<br>• Validity | • Precisely defined and specified<br>• Reliable<br>• Valid<br>• Risk-adjusted or stratified | • Clinical evidence linking processes, outcomes, interventions<br>• Reproducibility<br>• Validity (face, construct, content)<br>• Accuracy<br>• Case-mix risk adjustment methods<br>• Comparability of data sources |

| Evaluation criteria and requirements | National Quality Forum | AHRQ Quality Indicators | National Healthcare Quality Report | The Joint Commission | National Committee for Quality Assurance |
|---|---|---|---|---|---|
| ***3. Usability*** | **Usable** | **Application** | | | |
| • Effective (understandable and clear) presentation and dissemination strategies exist<br>• Statistical testing can be applied to communicate when differences in performance levels are greater than would be expected by chance<br>• Has been used effectively in the past and/or has high potential for working well with other indicators currently in use<br>• Compelling content for stakeholder decision making | • Measure can be used by stakeholders for decision making<br>• Performance differences are statistically meaningful<br>• Performance differences are practically and clinically meaningful<br>• Risk stratification, risk adjustment, and other forms of recommended analyses can be applied appropriately<br>• Effective presentation and dissemination strategies exist<br>• Information produced can be used by at least one health care stakeholder audience to make a decision or take action<br>• Information about specific conditions under which the measure is appropriate to use has been given<br>• Methods to aggregate the measure with related measures are defined if determined to be more understandable and useful | • Have been used effectively in the past<br>• Have high potential for working well with other indicators currently in use | | • Can be interpreted and useful in the accreditation process | |

| | | | | | |
|---|---|---|---|---|---|
| ***4. Feasibility*** | **Feasibility** | | **Feasibility** | | **Feasibility** |
| • Consistent construction and assessment of the measure<br>• Feasible to calculate; benefits exceed financial and administrative burden of implementation<br>• Confidentiality concerns are addressed<br>• Audit strategy can be implemented, quality of data are known | • Data collection tied to care delivery when feasible<br>• Timing and frequency of measure collection are specified<br>• Benefit evaluated against financial and administrative burden of implementation<br>• Confidentiality concerns are addressed<br>• Audit strategy is designed and can be implemented | | • Availability of required data across the system<br>• Cost or burden of measurement<br>• Existence of prototypes<br>• Capacity of data and measure to support subgroup analyses | • Data collection effort is assessed | • Precise definition (under scientific soundness in other frameworks)<br>• Reasonable cost<br>• Logistical feasibility<br>• Confidentiality<br>• Auditability |

**Source:** Remus D, Fraser I. Guidance for using the AHRQ Quality Indicators for hospital-level public reporting or payment. Rockville, MD: Agency for Healthcare Research and Quality; 2004. AHRQ Pub. No. 04-0086-EF. Available at: http://qualityindicators.ahrq.gov/downloads/technical/qi_guidance_appendix_B.pdf.

# Appendix B

# Contacts for Chartered Value Exchanges

### California Chartered Value Exchange

**Cathie Markow**
Senior Manager, CCHRI, Pacific Business Group on Health
221 Main Street
San Francisco, CA 94105
(415) 615-6359
cmarkow@pbgh.org

**David S. Hopkins**
Director of Health Information Improvement, Pacific Business Group on Health
221 Main Street
San Francisco, CA 94105
(415) 615-6322
dhopkins@pbgh.org

### Colorado Value Exchange

**Donna Marshall**
Executive Director, Colorado Business Group on Health
200 South Sheridan Boulevard, Suite 105
Denver, CO 80226-8010
(303) 922-0939
donna.marshall@cbghealth.org
Assistant: Kristen Berg
(303) 922-0939
kristen.berg@cbghealth.org

### Connecticut – eHealthConnecticut, Inc.

**Scott Cleary**
Program Director
100 Roscommon Drive
Middletown, CT 06457
(860) 240-5600
scleary@gosmcpartners.com
www.ehealthconnecticut.org
Assistant: Shay Cully
(860) 240-5600
scully@gosmcpartners.com

### Indiana – Quality Health First Program

**J. Marc Overhage**
President and CEO
846 North Senate Avenue, Suite 300
Indianapolis, IN 46202
(317) 644-1750
marc.overhage@ihie.org
www.ihie.org
Assistant: Sandy Poremba
(317) 423-5579
sporemba@regenstrief.org

**Chris Schultz**
Program Director
846 North Senate Avenue, Suite 300
Indianapolis, IN 46202
(317) 644-1741
Chris.schultz@ihie.org
www.ihie.org
Assistant: Wanda Rudisell
(317) 644-1722
wrudisell@ihie.org
Jill LeMasters
(317) 644-1745
jlemasters@ihie.com

### Kentucky – Greater Louisville Value Exchange Partnership

**Karin Kennedy**
Program Director, Health Care Excel, Inc.
2629 Waterfront Parkway East Drive, Suite 200
Indianapolis, IN 46214
(317) 347-4568
kkennedy@hce.org
info@value-driven.org

### Louisiana Health Care Quality Forum
Cindy Munn
Executive Director
304 Laurel Street, Suite 2D
Baton Rouge, LA 70801
(225) 334-9299, Ext. 204
cmunn@lhcqf.org
www.lhcqf.org
Assistant: Lynda Gaiennie
(225) 334-9299
lgaiennie@lhcqf.org

### Maine Chartered Value Exchange Alliance
**Elizabeth Mitchell**
President and CEO, Maine Health Management Coalition
P.O. Box 357
22 Stonebrooke Road
Scarborough, ME 04070-0357
(207) 883-8141
emitchell@mehmc.org
www.mhmc.info

### Massachusetts Chartered Value Exchange
**Barbra Rabson**
Executive Director, Massachusetts Health Quality Partners
100 Talcott Avenue
Watertown, MA 02472
(617) 402-5015
brabson@mhqp.org
www.mhqp.org

**Micky Tripathi**
President and CEO, Massachusetts eHealth Collaborative
860 Winter Street
Waltham, MA 02451-1411
(781) 434-7905
mtripathi@maehc.org
www.maehc.org

### Michigan – Alliance for Health
**Lody Zwarensteyn**
President
1345 Monroe NW, Suite 256
Grand Rapids, MI 49505
(616) 248-3820
lody@afh.org
www.afh.org
Assistant: Linda Stouten, ls@afh.org

### Michigan – Greater Detroit Area Health Council
**Devorah Rich**
Project Director
333 West Fort Street, Suite 1230
Detroit, MI 48226
(313) 965-4134
drich@gdahc.org
www.gdahc.org
Assistant: Tiffany Brent
(313) 596-0818
slsd_assistant@gdahc.org

### Michigan Health Information Alliance
Thomas M. Smith Jr.
Project Director
2625 Denison Drive
Mt. Pleasant, MI 48858
(989) 293-8590
tmcnallyjr@aol.com
www.mihia.org

### Minnesota Healthcare Value Exchange
**Susan McDonald**
Health Care Purchasing Coordinator/Governor's Health Cabinet, Minnesota Department of Human Services
2550 University Avenue West, Suite 2
St. Paul, MN 55114
(651) 431-4870
susan.mcdonald@state.mn.us
Assistant: Christine Wasieleski
christine.m.wasieleski@state.mn.us

### Missouri/Kansas – Kansas City Quality Improvement Consortium

**Cathy Davis**
Co-Director, UAW-Ford Community Health Care Initiative
6000 N. Oak, Suite 300
Kansas City, MO 64118
(816) 453-4424
cdavis20@ford.com
www.kcqic.org

### Nevada Partnership for Value-Driven Health Care

**Deborah Huber**
Transparency Director, HealthInsight
6830 W. Oquendo Road, Suite 102
Las Vegas, NV 89118
(702) 933-7305
dhuber@healthinsight.org
Assistant: Amy Baldwin
(702) 933-7337
abaldwin@healthinsight.org

### New York Quality Alliance

**Betsy Mulvey**
Project Coordinator
90 State Street, Suite 825
Albany, NY 12207
(518) 462-2293
bmulvey@nyhpa.org
www.nyqa.org

### Ohio – Health Improvement Collaborative of Greater Cincinnati and HealthBridge

**Bob Graham**
Professor of Family Medicine
University of Cincinnati College of Medicine
Health Professions Building, Suite 141
Cincinnati, OH 45220
(513) 558-5004
graham3@ucmail.uc.edu

### Oregon Health Care Quality Corporation

**Nancy Clarke**
Executive Director
619 SW 11th Avenue, Suite 221
Portland, OR 97205
(503) 241-3571
nancy.clarke@q-corp.org
www.q-corp.org

### Pennsylvania – Aligning Forces for Quality-South Central PA

**Christine Amy**
Project Director
c/o Family First Health
116 S. George Street, 3rd Floor
York, PA 17401
(717) 801-4823
camy@wellspan.org
www.aligning4healthpa.org
Assistant: Joyce Ortiz
(717) 801-4830
jortiz@wellspan.org

### Pennsylvania – Pittsburgh Regional Health Initiative

**Karen Feinstein**
President and CEO
Centre City Tower, Suite 2400
650 Smithfield Street
Pittsburgh, PA 15222
(412) 594-2555
feinstein@jhf.org
www.prhi.org
Assistant: Millie Greene
(412) 594-2556
greene@jhf.org

### Tennessee – Healthy Memphis Common Table

**Renee Frazier**
Executive Director
1016 W. Poplar Avenue, Suite 106 (PMB 218)
Collierville, TN 38017-2687
(901) 861-8220
contact@healthymemphis.org
www.healthymemphis.org
Assistant: Jennifer Joslin
jennifer.joslin@healthymemphis.org
Project Director: Denise Bollheimer
(901) 448-7046
denise.bollheimer@utmg.org,

### Utah Partnership for Value-driven Health Care

**Christie North**
Strategy Director
348 East 4500 South, Suite 300
Salt Lake City, UT 84107
(801) 892-6613
cnorth@healthinsight.org

### Virginia Health Care Alliance

**Sallie S. Cook**
Chief Medical Officer
9830 Mayland Drive, Suite J
Richmond, Virginia 23233
(804) 289-5320, ext. 342
scook@vhqc.org
Assistant: Carol Copeland
(804) 289-5330
ccopeland@vhqc.org,

### Washington – Puget Sound Health Alliance

**Mary McWilliams**
Executive Director
2003 Western Avenue, Suite 600
Seattle, WA 98121
(206) 448-2570
mary@pugetsoundhealthalliance.org
www.pugetsoundhealthalliance.org

### Wisconsin Health Care Value Exchange

**Chris Queram**
President and CEO
P.O. Box 628578
Middleton, WI 53562
(608) 826-6837
cqueram@wchq.org
www.wchq.org
Executive Assistant:
Tammie R. Cawkins,
(608) 826-6841
tcawkins@wchq.org

# Appendix C

# Glossary

**Aligning Forces for Quality (AF4Q):** AF4Q, sponsored by the Robert Wood Johnson Foundation, helps participating communities achieve sustainable health care quality through multistakeholder local alliances. These alliances focus on three intersecting program areas: (1) developing local quality improvement resources to help health professionals improve care, (2) helping the public to use appropriate information in making health care decisions, and (3) working to increase performance measurement and public reporting of those measures. AF4Q communities actively promote nurse leadership in the effort and also focus on reducing disparities in care for racial and ethnic groups.

**Better Quality Information to Improve Care for Medicare Beneficiaries (BQI):** The BQI Project is a Centers for Medicare & Medicaid Services (CMS)-funded project in which the Delmarva Foundation for Medical Care subcontracted with six community collaboratives as pilot sites. These pilot sites tested methods to aggregate claims data from Medicare, commercial health plans, and, in some cases, Medicaid. The data were used to calculate and report quality measures for physician groups and, in some cases, individual physicians.

**Consumer Assessment of Healthcare Providers and Systems (CAHPS®):** Supported by AHRQ, the CAHPS program is a public-private initiative to develop standardized surveys of patients' experiences. The surveys cover ambulatory care (medical groups, individual clinicians, mental health providers, and health plans) and facility-level care (hospitals, nursing homes, and dialysis centers). Surveys consist of core and supplemental questions concerning access and wait times, patient-doctor communication, trust, continuity of care, coordination between primary care physicians and specialists, referrals, preventive care, experiences with office staff, and demographic characteristics. Health care organizations, public and private purchasers, consumers, and researchers use CAHPS results to assess patient-centered care, compare and report on performance, and improve quality of care. CAHPS survey results can be obtained through AHRQ's national CAHPS Benchmarking Database, the National Committee for Quality Assurance, the HQA Hospital Compare Web site, and other public report cards.

**Chartered Value Exchange (CVE):** A CVE is a multistakeholder community quality collaborative composed of public and private payers, health plans, providers, and consumers so designated and supported (in-kind) by the Department of Health and Human Services (HHS) and AHRQ. In aggregate, the 24 AHRQ CVEs involve 600 health care leaders and represent more than 124 million lives, or more than one-third of the U.S. population. CVEs are intended to increase transparency and accountability by providing public information about the cost and quality of health care; they benefit from peer-to-peer interaction supported by AHRQ's CVE Learning Network. CVE stakeholders drive the content of the Learning Network technical assistance, which includes the following eight areas:

1. Collaborative leadership and sustainability;
2. Public at-large engagement;
3. Quality and efficiency measurement;

4. Public reporting;
5. Provider incentives;
6. Consumer incentives;
7. Coordinated cross-organizational, cross-stakeholder quality improvement; and
8. Health information technology/health information exchange.

**Community quality collaboratives:** Community-based organizations of multiple stakeholders that might include health care providers, purchasers (employers, employer coalitions, Medicaid, and others), health plans, and consumer advocacy organizations that work together to improve health care at the local level. These collaboratives vary in level of sophistication and degree of organization.

**Denominator**: The lower part of the fraction used to calculate a rate or ratio using the total population of interest.

**Diagnostic Cost Groups (DCGs):** Diagnostic Cost Groups classify diagnoses from administrative data into clinical groupings to create an aggregated measure of expected resource use. The measure, called a “relative risk score,” is calculated at the individual patient level and quantifies the financial implications of the patient’s total “illness burden” or morbidity.

**Diagnosis-Related Groups (DRGs):** Developed for Medicare as part of the Inpatient Prospective Payment System, DRGs classify hospital discharges into one of several hundred groups expected to have similar hospital resource use. DRGs are assigned by an annually updated “grouper” program based on ICD-9-CM (defined below) diagnoses, procedures, age, sex, discharge status, and presence of complications or comorbidities. In the current version, known as Medicare Severity DRGs or MS-DRGs, cases with a single diagnosis or major operating room procedure are often classified into one of three mutually exclusive severity levels: with major comorbidity or complication (CC), with CC, or without CC.

**Electronic health record (EHR):** In health informatics, an electronic medical record (EMR) is considered to be one of several types of EHRs, but EMR and EHR are also used interchangeably. EHRs are sometimes defined as including other systems that keep track of medical information, such as practice management software that facilitates the day-to-day operations of a medical practice. Such software frequently allows users to capture patient demographics, schedule appointments, maintain lists of insurance payers, perform billing tasks, and generate reports.

**Healthcare Effectiveness Data and Information Set (HEDIS):** HEDIS is a set of standardized measures designed by the National Committee for Quality Assurance to evaluate the quality of health care and service provided by health plans and physicians.

**Healthcare Cost and Utilization Project (HCUP):** HCUP is a family of health care databases and related software tools and products developed through a Federal-State-industry partnership and sponsored by AHRQ. HCUP databases bring together the data collection efforts of State data organizations, hospital associations, private data organizations, and the Federal Government to create a national information resource of service-level health care data (HCUP involves a number of partners).

**Health information exchanges (HIEs):** HIEs provide the capability to move clinical information electronically between disparate health care information systems while maintaining the meaning of the information being exchanged. The goal of HIEs is to facilitate access to and retrieval of clinical data to provide safer, more timely, efficient, effective, equitable, patient-centered care. HIEs also provide the infrastructure for secondary use of clinical data for purposes such as public health; clinical, biomedical, and consumer health informatics research; and institution and provider quality assessment and improvement.

**Health information technology (HIT):** HIT provides the umbrella framework to describe the comprehensive management of health information and its secure exchange between consumers, providers, government, quality entities, and insurers. HIT in general is viewed as a promising tool for improving the overall quality, safety and efficiency of the health delivery system.

**ICD-9-CM:** International Classification of Diseases, Clinical Modification (ICD-9-CM) is a classification used to assign codes to diagnoses associated with hospital, laboratory, and physician office utilization in the United States. The ICD-9-CM is based on the World Health Organization's ICD-9 but provides additional morbidity detail and is annually updated. In addition, the Cooperating Parties that created ICD-9-CM added a volume with procedure codes.

**Institute of Medicine (IOM):** The Institute of Medicine (IOM) is a not-for-profit, nongovernmental organization chartered in 1970 as part of the National Academy of Sciences. Its purpose is to provide national advice on issues related to biomedical science, medicine, and health, and its mission is to serve as adviser to the Nation to improve health. IOM provides independent guidance and evidence-based analysis, relying on a volunteer workforce of scientists and other experts operating under a rigorous, formal peer-review system.

**The Joint Commission (TJC):** The Joint Commission, formerly the Joint Commission on Accreditation of Healthcare Organizations, is a private nonprofit organization that evaluates health care facility compliance with Federal and industry standards through its accreditation process. The declared mission of this organization is "to continuously improve the safety and quality of care provided to the public through the provision of health care accreditation and related services that support performance improvement in health care organizations."

**Leapfrog Group:** Leapfrog is an employer-based group that seeks to "trigger giant leaps forward in the safety, quality and affordability of health care." Leapfrog surveys hospitals and scores their adherence to 13 of the 27 safe practice areas identified by the National Quality Forum. The Leapfrog survey also includes items about adherence to evidence-based processes of care for several high-risk conditions and procedures. The results are posted at www.leapfroggroup.org/ for public use.

**National Health Information Network (NHIN):** The U.S. Department of Health and Human Services awarded contracts to four groups of health care and health information technology organizations to develop prototypes for NHIN architecture. The contracts awarded to these four consortia are designed to move the Nation toward the President's goal of personal electronic health records by creating uniform architecture for health care information that can follow consumers throughout their lives.

**Mean:** In statistics, the mean is the mathematical average of a set of numbers. The mean is calculated by adding two or more scores and dividing the total by the number of scores.

**Median:** A median is described as the number separating the higher half of a sample from the lower half. The median of a finite list of numbers can be found by arranging all the observations from lowest value to highest value and picking the middle one. If there is an even number of observations, the median is not unique, so one often takes the mean of the two middle values.

**Medstat Medical Episode Groups® (MEGs):** Medstat's MEG is a commercial software package that enables health plans to analyze patient treatments, evaluate quality of care, and manage associated costs. It does so by grouping inpatient, outpatient, and pharmaceutical claims into clinically homogeneous units of analysis called "episodes." Each episode describes a patient's complete course of care for a single illness or condition. Results are used for provider profiling, disease management, quality improvement, and cost and utilization analyses.

**Numerator**: The upper part of the fraction used to calculate a rate or ratio using a subset of a population of interest (e.g., those patients with a heart attack who received bypass surgery).

**Pay-for-performance (P4P) programs:** P4P is a strategy to improve health care delivery that relies on the use of market or purchaser power. Depending on the context, P4P refers to financial incentives that reward providers for the achievement of a range of payer objectives, including delivery efficiencies, submission of data and measures to a payer, and improved quality and patient safety.

**Physician thumbprint:** A term used to describe individual physician effect on quality of care. It can be expressed as a composite measure of various metrics influenced more by physicians than patients.

**Quality Improvement Organization (QIO):** QIOs are private organizations that implement improvements in the quality of care for their State or region and respond to quality-related complaints from Medicare beneficiaries. By law, the mission of the QIO Program is to improve the effectiveness, efficiency, and quality of services delivered to Medicare beneficiaries. Based on this statutory charge, CMS identifies the core functions of the QIO Program as improving quality of care for beneficiaries and protecting the integrity of the Medicare Trust Fund. QIOs help ensure that Medicare pays only for services and goods that are reasonable and necessary and that are provided in the most appropriate setting. QIO responsibilities also include addressing reported violations of the Emergency Medical Treatment and Labor Act (EMTALA) and other related responsibilities as articulated in Federal law.

**Regional Health Information Organizations (RHIO):** RHIOs are multistakeholder organizations responsible for integrating health information exchange in the United States. RHIOs seek to affect the safety, quality, and efficiency of health care as well as access to health care through health information technology. RHIOs are a specialization of HIE.

**Risk adjustment:** Risk adjustment is a statistical process used to identify and adjust for variation in patient outcomes that stem from differences in patient characteristics (or risk factors) across health care organizations. Depending on the presence of risk factors at the time of health care encounters, patients may experience different outcomes regardless of the quality of care provided by the health care organization. Comparing patient outcomes across organizations without appropriate risk adjustment can be misleading. By adjusting for the risks associated with outcomes of interest, risk adjustment facilitates a more fair and accurate interorganizational comparison.

**Relative Value Units (RVUs):** Medicare uses a physician fee schedule to determine payments for more than 7,000 physician services. The fee for each service depends on its relative value units (RVUs), which rank on a common scale the resources used to provide each service. These resources include the physician's work, expenses of the physician's practice, and professional liability insurance. To determine the Medicare fee, a service's RVUs are multiplied by a dollar conversion factor. Estimating and updating the RVUs is a labor-intensive process because there are no readily available, up-to-date data on the resource requirements of each service.

**Symmetry Episode Treatment Groups (ETGs):** Symmetry Episode Treatment Groups® (ETGs®) is another example of commercial software that classifies health conditions and groups related treatment by episode of care.

**Taxpayer Identification Number (TIN):** A TIN is an identification number used by the Internal Revenue Service (IRS) in the administration of tax laws. It is issued either by the Social Security Administration or by the IRS. As the financial entity that receives payment on physicians' behalf, it does not necessarily correspond to a physician practice site or even a physician organization.

**Unique Physician Identification Number (UPIN):** The UPIN is a unique number for each physician and limited licensed practitioner who is enrolled in the Medicare program. Each UPIN can be associated with a practitioner's full name, specialty, license number, ZIP Code, and State.