

# **Refinement of the HCUP Quality Indicators**



Number 4

## **Refinement of the HCUP Quality Indicators**

**Prepared for:**

Agency for Healthcare Research and Quality  
U.S. Department of Health and Human Services  
2101 East Jefferson Street  
Rockville, MD 20852  
[www.ahrq.gov](http://www.ahrq.gov)

**Contract No. 290-97-0013**

**Prepared by:**

UCSF-Stanford Evidence-based Practice Center

Sheryl M. Davies, M.A.  
Jeffrey Geppert, J.D.  
Mark McClellan, M.D., Ph.D.  
Kathryn M. McDonald, M.M.  
Patrick S. Romano M.D., M.P.H  
Kaveh G. Shojania, M.D.  
*Core Project Team*

**AHRQ Publication No. 01-0035**  
**May 2001**

**Contributors**

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

**Consultant**

Douglas Staiger, Ph.D.

**AHRQ Contributors**

Anne Elixhauser, Ph.D.  
Margaret Coopey, R.N., M.G.A, M.P.S.

---

**Disclaimer**

The authors of this report are responsible for its content. Statements in the report should not be construed as endorsement by the Agency for Healthcare Research and Quality or the U.S. Department of Health and Human Services of a particular drug, device, test, treatment, or other clinical service.

The Agency does not guarantee the accuracy of this report. Questions regarding the content of this report, including all tables, figures, copyrights, and reference citations must be directed to the Evidence-based Practice Center that developed the report.

**U.S. Department of  
Health and Human Services**

Public Health Service  
Agency for Healthcare Research and Quality  
2101 East Jefferson Street, Suite 501  
Rockville, MD 20852



[www.ahrq.gov](http://www.ahrq.gov)

AHRQ Pub. No. 01-0035

May 2001

ISBN 1-58763-043-5

# Executive Summary

## Introduction

Healthcare quality has received heightened attention over the last decade, leading to a growing demand by providers, payers, policy makers, and patients for information on quality of care to help guide their decisions and efforts to improve health care delivery. At the same time, progress in electronic data collection and storage has enhanced opportunities to provide data related to health care quality. In 1989, the Agency for Health Care Policy and Research (AHCPR, now the Agency for Healthcare Research and Quality, AHRQ) initiated the Healthcare Cost and Utilization Project (HCUP). HCUP is an ongoing federal-state-private collaboration to build uniform databases from administrative hospital-based data collected by state data organizations and hospital associations. The first products of the collaboration were: 1.) creation of a comprehensive dataset of inpatient administrative records called the HCUP Nationwide Inpatient Sample (NIS), and 2.) development of a set of healthcare quality indicators (QIs).

The HCUP quality indicator set, developed in 1994, and hereafter referred to as HCUP I, consists of 33 measures, constructed using administrative data available in the NIS. Included in the set are indicators of utilization of procedures, ambulatory care sensitive condition admissions, post-operative and other complications, and mortality. Many measurement systems rely on extensive and expensive data collection, causing financial burdens on health care organizations and making ongoing and comprehensive monitoring of quality of care less likely. The HCUP indicators were developed as a low-cost, ongoing quality measurement mechanism for states able to develop standardized hospital discharge data. Due to the limitations of such administrative data, the indicators were intended for use as a screening tool rather than an absolute measurement of quality problems. Primarily, these indicators were based on measures described in the literature at the time of development. Further, the indicators were defined to be empirically simple; broad “denominator” populations were used in lieu of complicated risk adjustment systems.

Since the original HCUP QI development work in 1994, numerous managed care organizations, state Medicaid agencies and hospital associations, quality improvement organizations, the Joint Commission for the Accreditation of Healthcare Organizations (JCAHO), the National Committee on Quality Assurance (NCQA), academic researchers and others have contributed substantially to the knowledge base of hospital quality indicators. Based on input from current users and advances to the scientific base for specific indicators, AHRQ decided to fund a research project to refine and further develop the HCUP QIs. As a result, AHRQ charged the UCSF-Stanford Evidence-based Practice Center (EPC) to revisit the initial 33 indicator set (HCUP I QIs), evaluate their effectiveness as indicators, identify potential new indicators, and ultimately propose a revised set of indicators. This report documents the evidence project to develop recommendations for improvements to the HCUP I indicators.

In evaluating potential quality indicators, we applied the Institute of Medicine’s widely cited definition of quality of care 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.” We further focused on the clinical domains of potential underuse, overuse, and misuse, and excluded potential indicators based on patient satisfaction, health professional satisfaction, or cost containment. Only indicators ascertainable from current HCUP data were eligible for detailed review and empirical analysis. This report also excludes indicators relating to potential complications of care, because this set will be included in a separate evidence report covering patient safety indicators.

Three primary goals were established to accomplish this task:

- 1) Identify indicators in use and potential indicators.
- 2) Evaluate existing HCUP indicators and potential indicators using both literature review and empirical analyses of indicator performance.
- 3) Examine the need for risk adjustment of recommended indicators.

The team designed a series of investigations to accomplish these goals. These included telephone interviews of a small, purposeful sample of individuals knowledgeable about quality measurement, two phases of extensive literature reviews, and a series of empirical analyses using the State Inpatient Data (SID) data sets from 5 states. The in-depth review, supplemented by extensive empirical evaluation, focused on information that would be useful for implementing a revised set of HCUP quality indicators.

## Reporting the Evidence

The approach to identification and evaluation of QIs presented in this report serves as the basis for development of the revised HCUP QIs, hereafter referred to as HCUP II. The primary goal of the report is to document the evidence, both from the literature and from empirical analysis, on quality indicators suitable for use based on hospital discharge abstract data. By identifying and evaluating potential indicators, the report may serve as a springboard for commentary on proposed recommendations for specific improvements to the HCUP I QIs.

Six specific key questions were formulated to guide the research process:

- What indicators are currently in use or described in the literature that could be defined using HCUP discharge data?
- What are the quality relationships reported in the literature that could be used to define new indicators using HCUP discharge data?
- What evidence exists for indicators in AHRQ’s designated expansion areas – pediatric conditions, chronic disease, new technologies, and ambulatory care sensitive conditions?
- Of the existing HCUP I and potential indicators, which ones have literature-based evidence to support face validity, precision of measurement, minimum bias, and construct validity of the indicator?
- What risk-adjustment method should be supported, given the limits of administrative data and other practical concerns, for use in conjunction with the recommended indicators?

- Of the existing HCUP I and potential indicators, which ones perform well on empirical tests of precision of measurement, minimum bias, and construct validity?

The results of this project are 1) this evidence report, that summarizes all analyses and evaluations, and 2) software that can be used with hospital discharge data such as HCUP data (written in SAS™ programming language).

## Methodology

### Interviews

The project team interviewed a purposeful sample of 31 quality measurement stakeholders and experts affiliated with hospital associations, business coalitions, state data groups, federal agencies, and academia. These individuals, most of whom were either current or prospective users of HCUP QIs, provided the project team with background information regarding quality indicator use, suggested new indicators and risk adjustment methods, and helped frame our evaluation of potential indicators. (Interview methods are described in detail in Section 2.A. of the full report).

### Development of Evaluation Framework

Based on the interviews and a review of the relevant literature, the project team developed an *evaluation framework* of ideal standards by which to judge quality indicator performance:

- *Face validity.* An adequate quality indicator must have sound clinical and or empirical rationale for its use. It should measure an important aspect of quality that is subject to provider or health care system control.
- *Precision.* An adequate quality indicator should have relatively large variation among providers that is not due to random variation or patient characteristics.
- *Minimum bias.* The indicator should not be affected by systematic differences in patient case-mix, including disease severity and comorbidity. In cases where such systematic differences exist, an adequate risk adjustment system should be available based on HCUP discharge data.
- *Construct validity.* The indicator should be supported by evidence of a relationship to quality, and should be related to other indicators intended to measure the same or related aspects of quality.
- *Fosters Real Quality Improvement.* The indicator should not create incentives or rewards for providers to improve measured performance without truly improving quality of care.
- *Application.* The indicator should have been used effectively in the past, and/or have high potential for working well with other indicators currently in use.

In applying these criteria, the research team also considered the completeness of the evidence: obviously, it was more difficult to reach conclusions about each of these topics for indicators that had not been evaluated much in previous research. (More detail regarding the evaluation framework is available in Section 2.B. of the full report).

## Literature review

The literature review was completed in two phases. The first phase was designed to identify potential indicators. Quality indicators could be applicable to comparisons among providers of health care (e.g., hospitals, health systems) or among geographic areas (e.g., metropolitan service areas, counties), and should be applicable to a majority of providers or areas (i.e. not highly specialized care such as burn units). The second phase included a detailed review of the evidence on each indicator identified in Phase 1 using the criteria described in our evaluation framework. (Figure 1S diagrams the literature review process. Literature methods are described in detail in Section 2.C. of the full report).

*Phase 1.* To identify potential indicators, we performed a structured review of the literature. Using Medline, we identified the search strategy that returned a test set of known applicable articles in the most concise manner. The final MeSH terms used were “hospital, statistic and methods” and “quality indicators.” This search resulted in over 2000 articles published during or since 1994. These articles were screened for relevancy to this project according to specified criteria. The yield from the search and screen was 181 relevant articles.

Information from these articles was abstracted in two stages by clinicians, health services researchers and other team members. The first stage, preliminary abstraction, involved evaluation of each of the 181 identified articles for the presence of a defined quality indicator, potential quality indicators, and obvious strengths and weaknesses. To qualify for full abstraction (stage 2 of phase 1), the articles must have explicitly defined and evaluated a novel quality indicator. Similar to previous attempts to cull new indicators from the peer reviewed literature, few articles (27) met this criterion. Information on the definition of the quality indicator, validation and rationale were collected during full abstraction.

Additional potential indicators were identified using the CONQUEST (Computerized Needs-oriented Quality Measurement Evaluation SysTem) database, a list of ORYX™ approved indicators provided by the Joint Commission on Accreditation of Healthcare Organizations (JCAHO), Healthy People 2010 reports, and from the interviews and known web sites.

*Phase 2.* The inventory assembled from Phase 1 consisted of over 200 potential indicators. Initially, team members evaluated the clinical rationale of each indicator, and selected the most promising indicators based on a preliminary evaluation according to certain criteria, including minimum frequency of the event and sound clinical rationale. HCUP I indicators were not evaluated in this stage; they were automatically selected for the next step of evaluation. Second, indicators passing the initial screen (including the HCUP I indicators) were evaluated according to basic empirical tests of precision, including significant variation across providers, as described below. Third, a full literature review was conducted for those indicators with adequate performance on empirical precision tests. Medline was searched for articles relating to each of the six areas of evaluation, described in the evaluation framework. Clinicians, health services researchers and other team members searched the literature for evidence, and prepared a

referenced summary description of the evidence from the literature on each indicator. Each of these indicators also underwent a full empirical evaluation (see below).

## **Risk adjustment review and selection**

The literature regarding risk adjustment systems was reviewed. Alternative adjustment approaches for each indicator described in the literature were examined according to type of indicator (mortality, utilization, volume, ambulatory care sensitive condition) and analytic approach, method of development, feasibility of implementation given data availability, and empirical measures of discrimination and calibration. The evidence from the literature and information collected in the interviews with potential HCUP users were used to identify a practical method for risk adjustment of HCUP indicators.

Few risk adjustment systems could be feasibly implemented, given the lack of ambulatory, clinical, and longitudinal patient information in the current HCUP database. Diagnosis Related Group (DRG) systems fit more of the user preference-based criteria than other alternatives. In particular, a majority of users interviewed already used All Patients Refined (APR)-DRGs, and APR-DRGs have been reported to perform equivalently or better in predicting resource use and death for most indicators, when compared to other DRG based systems. Where feasible, the APR-DRG system was used to determine the effect of risk adjustment on the measured performance of providers on each reviewed indicator. (Risk adjustment methods are described in detail in Section 2.D. of the main report).

## **Empirical methods**

Extensive empirical testing of all potential indicators was conducted (See Tables 1-3 for a summary of empirical tests). In this overview, we provide a summary of the data sets used, and the specific tests for each of the evaluation criteria that were assessed empirically: precision, bias, and construct validity.

*Data set.* The primary data sets used were the HCUP Nationwide Inpatient Sample and the State Inpatient Database for 1995-1997. The annual NIS consists of about 6,000,000 discharges and over 900 hospitals from participating states. The SID contains all discharges for the included states. Most of the statistical tests used to compare candidate indicators were calculated using the SID, because the provider level results were similar to the NIS, and the SID includes all discharges for the calculation of area rates.

*Precision.* The first step in the analysis involved precision tests to determine the reliability of the indicator for distinguishing real differences in provider performance. Any quality indicator consists of both signal ('true' quality, that is what is intended to be measured) and noise (error in measurement due to sampling variation or other non-persistent factors). For indicators that may be used for quality improvement or other purposes, it is important to know with what precision, or surety, a measure can be attributed to an actual construct rather than random variation. For some indicators, the precision will be quite high for the raw measure. For other indicators, the precision will be rather low. However, it is possible to apply additional statistical techniques to improve the precision of these indicators. These techniques are called signal extraction, and are designed to "clean" or "smooth" the data of noise, and extract the actual signal associated with provider or area performance. We used two techniques for signal extraction to potentially improve the precision of an indicator. Detailed methods are contained in the methods section of the main report (Section 2.C). First, univariate methods estimated the

“true” quality signal of an indicator based on information from the specific indicator and one year of data. Second, new multivariate signal extraction (MSX) methods estimated the signal based on information from a set of indicators and multiple years of data. In most cases, MSX methods extract additional signal.

*Bias.* To provide empirical evidence on the sensitivity of candidate QIs to potential bias from differences in patient severity, we compared unadjusted performance measures for specific hospitals with performance measures that were adjusted for age, gender, and, where possible, patient clinical factors available in discharge data. We used the 3M APR-DRG System Version 12 with Severity of Illness and Risk of Mortality subclasses, as appropriate, for risk adjustment of the hospital quality indicators. For a few measures, no APR-DRG severity categories were available, so that unadjusted measures were compared to age-sex adjusted measures. Because HCUP data do not permit the construction of area measures of differences in risk, only age-sex adjustment is generally feasible for area-level indicators. We used a range of bias performance measures, most of which have been applied in previous studies. We note that these comparisons are based entirely on discharge data. In general, we expect performance measures that are more sensitive to risk adjustment using discharge data also to be more sensitive to risk adjustment using more complete clinical data, though the differences between the adjusted and unadjusted measures may be larger in absolute magnitude than the discharge data analysis would suggest. However, there may not be a correlation between discharge and clinical-record adjustment. Specific cases where previous studies suggest a greater need for clinical risk adjustment are discussed in our literature reviews of relevant indicators. To investigate the degree of bias in a measure, we performed five empirical tests (Spearman rank correlation, percentage remaining in extreme deciles, absolute change, percentage changing more than 2 deciles). Each test was repeated for the “raw” data, for data smoothed by univariate techniques (one year of data, one indicator), and for data smoothed by multivariate (MSX) techniques (using multiple years of data, all indicators).

*Construct validity.* Two measures of the same construct would be expected to yield similar results. If quality indicators do indeed measure quality, at least in a specified domain, such as ambulatory care, one would expect measures to be related. As quality relationships are likely to be complex, and outcomes of medical care are not entirely explained by quality, perfect relationships between indicators seem unlikely. We performed analyses to assess the potential relationships between indicators. To measure the degree of relatedness between indicators, we conducted a factor analysis, a statistical technique used to reveal underlying patterns among large numbers of variables. The output for a factor analysis is a series of “factors” or overarching constructs, for which each indicator would “load” or have a relationship with others in the same factor. The assumption is that indicators loading strongly on the same factor are related to each other via some independent construct. We used an orthogonal rotation to maximize the possibility that each indicator would load on one factor only, to ease the interpretation of the results. In addition to the factor analysis, we also analyzed correlation matrices for each type of indicator (provider level, ambulatory care sensitive condition (ACSC) area level, and utilization area level).

The construct validity analyses provided information regarding the relatedness or independence of the indicators. Such analyses cannot prove that quality relationships

exist, but they can provide preliminary evidence on whether the indicators appear to provide consistent evidence related to quality of care. For hospital volume quality indicators, we evaluated correlations with other volume and hospital mortality indicators, to determine whether the proposed HCUP II indicators suggested the same types of volume-outcome relationships as have been demonstrated in the literature.

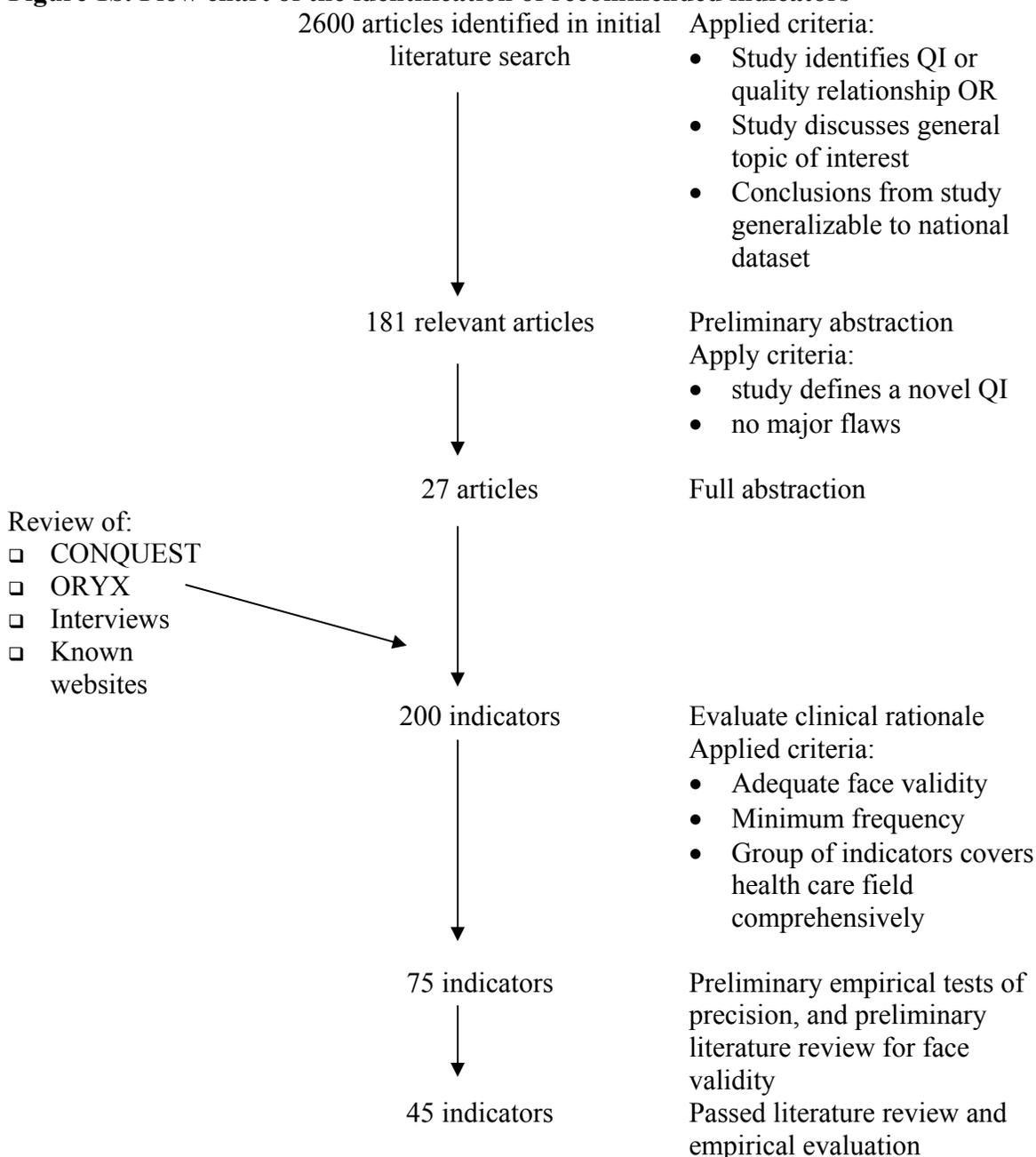
*Results of empirical evaluations.* Statistical test results for candidate indicators were compared. First, the results from precision tests were used to sort the indicators. Those indicators performing poorly were eliminated. Second, the results from bias tests were conducted to determine the need for risk adjustment. Finally, construct validity was determined to provide some evidence on the nature of the relationship between potential indicators.

## Results

Over 200 indicators (listed in Appendix 7 of the full report) that could be specified using inpatient discharge data, such as the HCUP NIS, and that met our criteria for “quality indicator,” (i.e. examined an aspect of quality as defined above, applicable to most providers/areas) were identified and evaluated as potential HCUP QIs. Based on our preliminary application of criteria for indicator validity, 45 promising indicators were retained for comprehensive literature and empirical evaluation. In some cases, whether an indicator complemented other promising indicators was a consideration in retaining it, allowing the HCUP indicators to provide more depth in specific areas.

The Evidence Report provides detailed literature summaries and data from empirical analyses on each of the 45 indicators. The indicators were constructed, as appropriate, for two perspectives on quality — “provider-level” and “area-level”. Provider-level indicators are designed using a hospital-level denominator. Area-level indicators are designed with population-based denominators, specifically the population of the metropolitan statistical area (MSA). There are 25 provider-level quality indicators and 20 area-level indicators recommended for use.

While none of these indicators is without its limitations, a considerable literature in most cases coupled with evidence on satisfactory empirical performance suggests that the recommended indicators may be useful additions to the “toolkit” for clinical quality professionals, health care managers, health policymakers, as well as researchers. Each of the recommended indicators is appropriate for use as a quality “screen,” or as a first examination of potential quality problems, to be followed up by more in-depth investigations. Our evaluation noted the most promising uses of each indicator, as well as important limitations and suggestions for further investigation.

**Figure 1S. Flow chart of the identification of recommended indicators****Provider indicators**

Provider indicators are constructed at the provider level; they provide information related to the quality of care at individual hospitals. There are four types:

- *Volume indicators* include inpatient procedures for which a substantial research literature has detected a significant relationship between hospital volume and outcomes, and for which a nontrivial number of procedures are performed by institutions that do not meet recommended volume thresholds. The volume indicators are somewhat different than the other provider-level indicators, in that they simply

represent counts of admissions in which particular intensive procedures were performed rather than more direct measures of quality.

- *Utilization indicators* include procedures whose use varies significantly across hospitals, and for which high or low rates of use are likely to represent inappropriate or inefficient delivery of care, leading to worse outcomes, higher costs or both.
- *Mortality indicators for inpatient procedures* include those for which mortality has been shown to vary substantially across institutions and for which evidence suggests that high mortality may be associated with deficiencies in the quality of care.
- *Mortality indicators for inpatient conditions* include those for which mortality has also been shown to vary substantially across institutions, and for which evidence suggests that high mortality may be associated with deficiencies in the quality of care.

### **Area Indicators**

The evidence report includes a set of quality indicators constructed at the area level. Versions of some of these indicators were previously recommended as HCUP I indicators. However, their construction differs in that the denominator for the indicators is now constructed at the area level. For most of these indicators, the denominator is the age- and gender-adjusted *population*, and the numerator is the rate of hospitalization with the procedure or diagnosis. These indicators are constructed at the level of metropolitan statistical areas (MSA). At the county level (a finer area measure), evidence from Medicare and California data suggest that a significant proportion of patients at many hospitals come from outside the area, and many patients from an area seek care at facilities in other areas. At the MSA level, the vast majority of patients treated in an MSA come from the MSA; and the vast majority of residents of an MSA receive treatment in the MSA. With more detailed information on patient residence (not available currently in the HCUP NIS), richer and more accurate area indicators could be constructed using the definitions applied in this report. There are two types of area indicators assessed:

- *Utilization indicators* include procedures for which use has been shown to vary widely across relatively similar geographic areas, with (in most cases) substantial inappropriate utilization.
- *Avoidable hospitalizations/ Ambulatory care sensitive condition (ACSC) indicators* involve admissions that evidence suggests could have been avoided, at least in part, through better access to high quality outpatient care.

Even though these quality indicators are area-based, an important role remains for hospital-level measures of procedures or ACSC admissions. If an area is found to have unusually high procedure rates, a natural focus for efforts to understand why rates are high and possibly to reduce them is the particular hospitals that perform a relatively large proportion of the area procedures. Similarly, if an area is found to have unusually high admission rates for potentially avoidable conditions, then the patient populations treated

by hospitals with a relatively large share of these admissions might be a good starting point for interventions to understand and reduce hospitalization rates.

### Using indicators as groups

All indicators in isolation provide a unidimensional and fairly limited picture of quality. As the results of this report indicate, many factors besides quality may contribute to provider or area performance on a single quality indicator, including random variation. However, consistent good or bad performance on several related indicators is more convincing evidence of a true underlying difference in performance, as it is more unlikely that such a pattern could arise from random events. Looking at groups of indicators together, therefore, is likely to provide a more complete picture of quality. While the HCUP indicators were not designed to be averaged or combined into an overall quality score, they do group together both by clinical domain and by aspects of care or outcome. For example, CABG mortality rates must be viewed in the context of CABG utilization and volume (i.e., grouping by clinical domain), since inappropriate utilization for less severe patients may increase provider volumes and decrease postoperative mortality. Mortality rates for major medical diagnoses should also be viewed together (i.e., grouping by outcome), because skill in caring for community-acquired pneumonia would be expected to carry over to diagnoses such as congestive heart failure. This report does not present findings on the validity of such groupings, although some, such as the ACSC indicators, have been examined extensively elsewhere.

### Indicator Performance

As noted, each potential indicator underwent extensive evaluations based on literature reviews and empirical analyses. Table 1S (provider-level indicators) and Table 2S (area-level indicators) list each indicator, describe its definition, rate its empirical performance, recommend a risk adjustment strategy, and note important caveats identified in the literature reviews.

**Empirical performance rating.** Our rating of empirical performance is a numerical rating that ranges from 0 – 26. This rating summarizes the performance on four empirical tests of precision (signal variance, provider/area-level share, signal ratio, and r-square), and five tests of minimum bias (rank correlation, top/bottom decile movement, absolute change, and change over 2 deciles). Because we were better able to conclusively measure the precision of an indicator than minimum bias (because available risk adjustment techniques were not clinically comprehensive, and thus may underestimate some bias), we weighted precision tests more than minimum bias tests. Each indicator was given a score of 0 – 4 based on its performance on the precision tests, relative to the other indicators, and based on specific cutoffs described in the main document. Likewise, each indicator was given a score of 0-2 on each of the bias tests. The empirical performance rating is the sum of those nine scores. The mean for the provider indicators was 9.7 (S.D. = 6.5). The mean for the area indicators was 16.2 (S.D. = 3.4). This reflects primarily the better precision of area measures relative to mortality measures. In cases where multivariate smoothing techniques improve the amount of variance that can be attributed to true differences in performance, it is noted that smoothing is recommended.

**Caveats from the literature review.** During the review of the literature we identified serious and potential caveats for each of the recommended indicators. These

caveats tended to follow general themes, and are summarized in the table below. When specific evidence was found demonstrating that the caveat applies to that indicator, that caveat is preceded by a checkmark. When no such evidence was located, but there is a strong theoretical basis or suggestive evidence that the caveat applies, a question mark precedes the caveat name in the table. The specific caveats are described below, along with potential remedies.

*Proxy indicator.* Some indicators do not specifically measure a patient outcome or a process measure of quality. Rather, some indicators measure an aspect of care that has been correlated with process measures of quality or patient outcomes. The validity of these indicators relies on the persistent and strong relationship between the measured phenomenon and actual quality. For example, provider volume has been correlated with better outcomes for numerous procedures, but volume, in the absence of these relationships, does not tell one anything about quality. Area utilization measures are another example of proxy indicators. High procedure rates do not necessarily imply overuse or inappropriate utilization; for some areas, higher rates may actually represent better care.

In cases where this concern is noted, continued research on the relationship validating the indicator (such as volume-outcome relationships) is required to ensure the validity of this indicator. These indicators are best used in conjunction with other indicators measuring similar aspects of clinical care, or when followed with more direct and in-depth investigations of quality.

*Selection bias.* Selection bias results when the cases with a condition or procedure ascertainable from HCUP data do not represent the universe of patients with that condition or procedure. As a result, the rate of an indicator based on HCUP data may differ from the true value in the population. This problem arises when a substantial percentage of care for a condition or procedure is provided in the outpatient setting, so the subset of inpatient cases may be unrepresentative. For example, laparoscopic cholecystectomy rates based on HCUP data may be biased because hospitals admit all patients who require open cholecystectomy, but only some patients scheduled for laparoscopic cholecystectomy. Similarly, patients with mild congestive heart failure may be admitted at some hospitals, but managed as outpatients elsewhere. A related problem is that inadequate or variable coding of key diagnoses may interfere with consistent ascertainment of cases, such as for vaginal births after cesarean delivery.

In cases where this concern is noted, examination of outpatient care or patients not admitted to the hospital (e.g., ER data) may help to improve indicator performance. Better risk-adjustment may help reduce selection bias for mortality indicators, which is attributable to variation in the threshold for admission.

*Information bias.* HCUP II QIs are based on information available in hospital discharge data sets, but some missing information may actually be important to evaluating the outcomes of hospital care. For instance, for some conditions, 30-day mortality has been shown to substantially exceed in-patient mortality. Without 30-day mortality data (ascertained from death certificates), hospitals that have short lengths of stay may appear to have better patient outcomes than other hospitals with equivalent 30-day mortality.

In cases where this concern is noted, examination of missing information, such as 30-day mortality, may help to improve indicator performance.

*Confounding bias.* Patient characteristics, such as disease severity, comorbidities, physiologic derangements, and functional status, may substantially affect performance on a measure, and may vary systematically across providers or areas. We are especially concerned about confounders that cannot be identified from HCUP data, such as physical examination, laboratory, radiographic, and functional abnormalities.

In cases where this concern is noted, adequate risk adjustment may help to improve indicator performance. In some cases, such risk-adjustment may require only the demographic and comorbidity data captured by APR-DRGs or similar systems. In other cases, detailed clinical data may be necessary for adequate risk-adjustment.

*Unclear construct validity.* Many indicators have not been examined extensively in the literature, although they are currently in use by various health care organizations. Problems with construct validity include: (1) uncertain or poor correlations with widely accepted process measures, and (2) uncertain or poor correlations with risk-adjusted outcome measures. Although these indicators have adequate face validity, they would benefit from further research to establish their relationship with quality care.

*Easily manipulated.* When quality indicators are instituted, they may create perverse incentives to improve performance on the quality indicator without actually improving quality. Dysfunctional organizational responses might include “cherry-picking” the easiest cases, “teaching to the test” by ignoring broader aspects of quality, “deception” through “upcoding” of comorbidities used in risk adjustment, and by being overcritical of quality measurement efforts. Providers may admit or perform procedures on less severe patients with dubious indications in order to inflate their volumes and improve apparent performance. Although very few of these perverse responses have been proven to occur, they are important theoretical concerns that should be monitored to ensure true quality improvement.

*Unclear benchmark.* Some indicators have clear goals for performance. Fewer deaths is always better; fewer low birth weight infants is ideal. However, for a few indicators, the numerator may include appropriate and unavoidable occurrences. When there is a base “right rate” of the indicator, either too low a rate or too high a rate may be a quality problem. For procedure utilization and ACSC admissions, too low a rate may indicate poor access to care or underuse of appropriate care. For these indicators, the “right rate” has not been established, so comparison with national, regional, or peer group means may be the best benchmark available.

**Table 1S. Provider indicator list**

<b>Provider Level Indicator Name</b>	<b>Type of Indicator</b>	<b>Indicator Description</b>	<b>Empirical Rating<sup>a</sup></b>	<b>Risk Adjustment Recommended</b>	<b>Caveats of use from the literature review<sup>b</sup></b>	<b>Better quality may be associated with:</b>
AAA repair volume (#1)	Volume	Raw volume compared to annual thresholds (20, 32 procedures)	Not applicable	Not applicable	➤ Proxy ? Easily manipulated	Higher rates
Carotid endarterectomy volume (#2)	Volume	Raw volume compared to annual thresholds (50, 101 procedures)	Not applicable	Not applicable	➤ Proxy ➤ Easily manipulated	Higher rates
CABG volume (#3)	Volume	Raw volume compared to annual thresholds (100, 200 procedures)	Not applicable	Not applicable	➤ Proxy ? Easily manipulated	Higher rates
Esophageal resection volume (#4)	Volume	Raw volume compared to annual thresholds (6, 7 procedures)	Not applicable	Not applicable	➤ Proxy ? Easily manipulated	Higher rates
Pancreatic resection volume (#5)	Volume	Raw volume compared to annual thresholds (10, 11 procedures)	Not applicable	Not applicable	➤ Proxy ? Easily manipulated	Higher rates
Pediatric heart surgery volume (#6)	Volume	Raw volume compared to annual thresholds (100 procedures)	Not applicable	Not applicable	➤ Proxy ? Easily manipulated	Higher rates
PTCA volume (#7)	Volume	Raw volume compared to annual thresholds (200, 400 procedures)	Not applicable	Not applicable	➤ Proxy ? Selection bias ➤ Easily manipulated	Higher rates
Cesarean section rate (#8)	Utilization	Number of cesarean sections per 100 deliveries	17	Age adjustment, and potentially supplemental (clinical data, linked to infant record, or linked to birth record).	? Confounding bias ? Unclear construct validity ? Unclear benchmark	Lower rates <sup>c</sup>
Incidental appendectomy among elderly rate (#9)	Utilization	Number of incidental appendectomies per 100 abdominal surgeries	13 Smoothing recommended	Age and sex	? Unclear construct validity ? Easily manipulated	Lower rates

<b>Provider Level Indicator Name</b>	<b>Type of Indicator</b>	<b>Indicator Description</b>	<b>Empirical Rating<sup>a</sup></b>	<b>Risk Adjustment Recommended</b>	<b>Caveats of use from the literature review<sup>b</sup></b>	<b>Better quality may be associated with:</b>
Bi-lateral cardiac catheterization rate (#10)	Utilization	Number of bilateral cath per 100 cardiac cath	25	None required	? Selection bias ? Unclear construct validity	Lower rates
VBAC rate (#11)	Utilization	Number of vaginal births per 100 deliveries in women with previous cesarean section	19	Age adjustment, and potentially supplemental (clinical data, linked to infant record, or linked to birth record).	➤ Selection bias ? Confounding bias ? Unclear construct validity ? Unclear benchmark	Higher rates <sup>c</sup>
Laparoscopic cholecystectomy (#12)	Utilization	Number of lap. cholecystectomies per 100 cholecystectomies.	20	Age and sex adjustment, and potentially supplemental clinical.	➤ Selection bias ➤ Confounding bias ? Unclear construct validity ➤ Easily manipulated ➤ Unclear benchmark	Higher rates <sup>c</sup>
AMI mortality (#33)	In-hospital Mortality	Number of deaths per 100 discharges for AMI	5 Smoothing recommended	APR-DRG	➤ Information bias ➤ Confounding bias	Lower rates
CHF mortality (#34)	In-hospital Mortality	Number of deaths per 100 discharges for CHF	6 Smoothing recommended	APR-DRG	➤ Selection bias ➤ Information bias ➤ Confounding bias	Lower rates
GI hemorrhage mortality (#35)	In-hospital Mortality	Number of deaths per 100 discharges for GI hemorrhage	5 Smoothing recommended	APR-DRG	➤ Confounding bias ? Unclear construct validity	Lower rates
Hip fracture mortality (#36)	In-hospital Mortality	Number of deaths per 100 discharges for hip fracture	10 Smoothing recommended	APR-DRG	? Information bias ➤ Confounding bias ? Unclear construct validity	Lower rates
Pneumonia mortality (#37)	In-hospital Mortality	Number of deaths per 100 discharges for pneumonia	7 Smoothing recommended	APR-DRG	➤ Selection bias ? Information bias ➤ Confounding bias	Lower rates
Stroke mortality (#38)	In-hospital Mortality	Number of deaths per 100 discharges for stroke	10 Smoothing recommended	APR-DRG	➤ Selection bias ? Information bias ➤ Confounding bias	Lower rates
AAA repair mortality	Post-	Number of deaths per 100	8	APR-DRG, though	➤ Confounding bias	Lower rates

<b>Provider Level Indicator Name</b>	<b>Type of Indicator</b>	<b>Indicator Description</b>	<b>Empirical Rating<sup>a</sup></b>	<b>Risk Adjustment Recommended</b>	<b>Caveats of use from the literature review<sup>b</sup></b>	<b>Better quality may be associated with:</b>
(#39)	procedural Mortality	AAA repairs	Smoothing recommended	impact may be impaired by skewed distribution.	? Unclear construct validity	
CABG mortality (#40)	Post-procedural Mortality	Number of deaths per 100 CABG procedures	5 Smoothing recommended	APR-DRG	? Selection bias ➤ Confounding bias ? Unclear construct validity ? Easily manipulated	Lower rates
Craniotomy mortality (#41)	Post-procedural Mortality	Number of deaths per 100 craniotomies	6 Smoothing recommended	APR-DRG	➤ Confounding bias ? Unclear construct validity	Lower rates
Esophageal resection mortality (#42)	Post-procedural Mortality	Number of deaths per 100 esophageal resections for cancer	8 Smoothing recommended	APR-DRG, though impact may be impaired by skewed distribution.	? Confounding bias ? Unclear construct validity	Lower rates
Hip replacement mortality (#43)	Post-procedural Mortality	Number of deaths per 100 hip replacements	3 Smoothing recommended	APR-DRG	? Selection bias ? Confounding bias ? Unclear construct validity	Lower rates
Pancreatic resection mortality (#44)	Post-procedural Mortality	Number of deaths per 100 pancreatic resections for cancer	5 Smoothing recommended	APR-DRG, though impact may be impaired by skewed distribution.	? Confounding bias ? Unclear construct validity	Lower rates
Pediatric heart surgery mortality (#45) <sup>d</sup>	Post-procedural Mortality	Number of deaths per 100 heart surgeries in patients under age 18 years	3 Smoothing recommended	APR-DRG	➤ Confounding bias ? Unclear construct validity Unclear benchmark	Lower rates

<sup>a</sup>Each indicator is rated from 0-26 on its empirical performance of precision and minimum bias with 0 indicating the lowest empirical rating (poor performance) and 26 indicating the highest performance.

<sup>b</sup>Each indicator was evaluated for seven caveats (proxy, selection bias, information bias, confounding bias, unclear construct validity, easily manipulated, unclear benchmark). A question mark preceding the caveat means that this is a theoretical or suggested concern. A checkmark means that this is a concern that has been demonstrated.

<sup>c</sup>For some indicators, very low or very high rates may indicate a potential quality problem. The direction listed is the direction for improvement given the current rates for these indicators.

<sup>d</sup>Pediatric heart surgery mortality is not recommended as a stand alone indicator, because of ample evidence for confounding bias in the absence of more sophisticated risk-adjustment. It is designed only for use with the corresponding volume measure, or with risk-adjustment methods such as those described in the detailed literature review.

**Table 2S. Area indicator list**

Area Level Indicator Name	Type of Indicator	Indicator Description	Empirical Rating <sup>a</sup>	Risk Adjustment Recommended	Literature review findings <sup>b</sup>	Better quality may be associated with:
CABG rate (#13) <sup>d</sup>	Utilization	Number of CABGs per 100,000 population	19	Age and sex	<ul style="list-style-type: none"> <li>➤ Proxy</li> <li>➤ Unclear construct validity</li> <li>➤ Unclear benchmark</li> </ul>	Lower rates <sup>c</sup>
Hysterectomy rate (#14)	Utilization	Number of hysterectomies per 100,000 population	22	Age and additional factors such as parity.	<ul style="list-style-type: none"> <li>➤ Proxy</li> <li>? Confounding bias</li> <li>➤ Unclear construct validity</li> <li>➤ Unclear benchmark</li> </ul>	Lower rates <sup>c</sup>
Laminectomy rate (#15)	Utilization	Number of laminectomies per 100,000 population	20	Age and sex	<ul style="list-style-type: none"> <li>➤ Proxy</li> <li>➤ Unclear construct validity</li> <li>➤ Unclear benchmark</li> </ul>	Lower rates <sup>c</sup>
PTCA rate (#16) <sup>d</sup>	Utilization	Number of PTCAs per 100,000 population	19	Age and sex	<ul style="list-style-type: none"> <li>➤ Proxy</li> <li>? Selection bias</li> <li>➤ Unclear construct validity</li> <li>➤ Unclear benchmark</li> </ul>	Lower rates <sup>c</sup>
Dehydration admission rate (#17) <sup>d</sup>	ACSC	Number of admissions for dehydration per 100,000 population	14	Age and sex	<ul style="list-style-type: none"> <li>? Proxy</li> <li>? Unclear construct validity</li> <li>? Easily manipulated</li> <li>➤ Unclear benchmark</li> </ul>	Lower rates <sup>c</sup>
Bacterial pneumonia admission rate (#18) <sup>d</sup>	ACSC	Number of admissions for bacterial pneumonia per 100,000 population	17	Age and sex	<ul style="list-style-type: none"> <li>? Proxy</li> <li>? Unclear construct validity</li> <li>? Easily manipulated</li> <li>➤ Unclear benchmark</li> </ul>	Lower rates <sup>c</sup>
Urinary infection admission rate (#19) <sup>d</sup>	ACSC	Number of admissions for urinary infection per 100,000 population	11	Age and sex	<ul style="list-style-type: none"> <li>? Proxy</li> <li>? Unclear construct validity</li> <li>? Easily manipulated</li> <li>➤ Unclear benchmark</li> </ul>	Lower rates <sup>c</sup>
Perforated appendix admission rate (#20)	ACSC	Number of admissions for perforated as a share of all admissions for appendicitis within an area	17 Smoothing recommended	Age and sex	? Proxy	Lower rates
Angina admission	ACSC	Number of admissions for	19	Age and sex	? Proxy	Lower rates <sup>c</sup>

Area Level Indicator Name	Type of Indicator	Indicator Description	Empirical Rating <sup>a</sup>	Risk Adjustment Recommended	Literature review findings <sup>b</sup>	Better quality may be associated with:
rate (#21) <sup>d</sup>		angina per 100,000 population			? Unclear construct validity ? Easily manipulated ➤ Unclear benchmark	
Adult asthma admission rate (#22)	ACSC	Number of admissions for asthma in adults per 100,000 population	16	Age and sex	? Proxy ? Easily manipulated ➤ Unclear benchmark	Lower rates <sup>c</sup>
COPD admission rate (#23)	ACSC	Number of admissions for COPD per 100,000 population	17	Age and sex. Potential supplement includes patient characteristics, such as smoking.	? Proxy ? Confounding bias ? Easily manipulated ➤ Unclear benchmark	Lower rates <sup>c</sup>
CHF admission rate (#24)	ACSC	Number of admissions for CHF per 100,000 population	14	Age and sex.	? Proxy ? Easily manipulated ➤ Unclear benchmark	Lower rates <sup>c</sup>
Diabetes short term complication admission rate (#25)	ACSC	Number of admissions for diabetes short term complications per 100,000 population	14	Age and sex. Potential supplement includes population diabetes incidence rates.	? Proxy ? Confounding bias	Lower rates
Uncontrolled diabetes admission rate (#26) <sup>d</sup>	ACSC	Number of admissions for uncontrolled diabetes per 100,000 population	14	Age and sex. Potential supplement includes population diabetes incidence rates.	? Proxy ? Confounding bias ? Easily manipulated	Lower rates
Diabetes long term complication admission rate (#27)	ACSC	Number of admissions for diabetes long term diabetes per 100,000 population	11	Age and sex. Potential supplement includes population diabetes incidence rates.	? Proxy ? Confounding bias ? Easily manipulated ➤ Unclear benchmark	Lower rates
Hypertension admission rate (#28)	ACSC	Number of admissions for hypertension per 100,000 population	14	Age and sex	? Proxy ? Easily manipulated ➤ Unclear benchmark	Lower rates <sup>c</sup>
Lower extremity	ACSC	Number of admissions for	10	Age and sex.	? Proxy	Lower rates

Area Level Indicator Name	Type of Indicator	Indicator Description	Empirical Rating <sup>a</sup>	Risk Adjustment Recommended	Literature review findings <sup>b</sup>	Better quality may be associated with:
amputation in diabetics rate (#29) <sup>d</sup>		lower extremity amputation in diabetics per 100,000 population	Smoothing recommended	Potential supplement includes population diabetes incidence rates.	? Unclear construct validity	
Low birth weight rate (#30) <sup>d</sup>	ACSC	Number low birth weight births as a share of all births in an area	11 out of 16 <sup>c</sup> Smoothing recommended	None available in HCUP. Potential supplement include clinical, link to mother's record, or link to birth record.	? Proxy ? Confounding bias ➤ Unclear construct validity	Lower rates
Pediatric asthma admission rate (#31)	ACSC	Number of admissions for pediatric asthma per 100,000 population	18	Age and sex	? Proxy ? Easily manipulated ➤ Unclear benchmark	Lower rates <sup>c</sup>
Pediatric gastroenteritis admission rate (#32) <sup>d</sup>	ACSC	Number of admissions for pediatric gastroenteritis per 100,000 population	17	Age and sex	? Proxy ? Unclear construct validity ? Easily manipulated ➤ Unclear benchmark	Lower rates <sup>c</sup>

<sup>a</sup>Each indicator is rated from 0-26 on its empirical performance of precision and minimum bias with 0 indicating the lowest empirical rating (poor performance) and 26 indicating the highest performance.

<sup>b</sup>Each indicator was evaluated for seven caveats (proxy, selection bias, information bias, confounding bias, unclear construct validity, easily manipulated, unclear benchmark). A question mark preceding the caveat means that this is a theoretical or suggested concern. A checkmark means that this is a concern that has been demonstrated.

<sup>c</sup>For some indicators, very low or very high rates may indicate a potential quality problem. The direction listed is the direction for improvement given the current rates for these indicators.

<sup>d</sup>CABG and PTCA area utilization are not recommended as stand alone indicators. They are designed only for use with the corresponding volume and/or mortality measures. Seven ACSC measures (low birth weight, angina, urinary tract infection, bacterial pneumonia, lower extremity amputation, pediatric gastroenteritis, and dehydration) do not have studies confirming their construct validity as individual indicators of access to health care services. For this reason, it is recommended that they be used in conjunction with other ACSC indicators. Uncontrolled diabetes is designed to be combined with diabetes short term complications.

<sup>e</sup>Adequate risk adjustment for low birth weight was not available, and thus, bias was not tested.

## Conclusions and Future Research

For use as screens for quality concerns, each of the indicators evaluated and included in this report performed adequately. In many cases, however, adequate performance required important statistical enhancements (risk adjustment, smoothing methods) beyond simply calculating average rates. These indicators, accompanied by statistical enhancements, are recommended for implementation into software modules to replace the current HCUP QI set. For users of these indicators, further investigations are likely to be necessary when an indicator flags a potential problem. That is, even if an indicator identifies “outlier” hospitals or areas with great degree of precision, the cause of systematic differences in performance may be something other than poor quality. Our report presents specific suggestions for such follow-up steps for each type of indicator; we summarize some of the general findings here.

### Provider level Volume Indicators

The HCUP QI empirical results confirm that hospital volume is an important correlate of quality of care. However, our empirical results as well as the prior studies summarized in the detailed reviews of each indicator also make clear that volume is at best a quite noisy reflection of true quality or performance differences. While hospital volume has significant explanatory power, the relationship is not precise; in practical terms, there appear to be many high-quality procedures performed by low-volume institutions, and conversely many low-quality procedures performed by high-volume institutions. Causes of the relatively weak relationship between volume and quality include the confounding role of surgeon volume (not captured presently in HCUP data), differences in the severity and complexity of cases treated, and differences in training and experience that are not reflected in volume. Moreover, use of volume as a quality indicator may lead to undesirable hospital responses, such as performing more procedures on patients who have mild disease or who are otherwise inappropriate candidates. Thus, while volume is a useful proxy for quality, it is important to consider more direct measures of hospital performance to help determine whether a high-volume hospital provides excellent quality of care, and whether a low-volume hospital provides poor quality of care.

### Provider level Mortality Indicators

The recommended hospital mortality indicators are all associated with large systematic differences in hospital performance, that is, differences in mortality outcomes between lower- and higher-performing hospitals are often several percentage points or larger. Thus, the mortality indicators may be helpful in identifying opportunities for large improvements in outcomes. However, many of the mortality indicators require careful attention to risk adjustment, and virtually all benefit from “smoothing” methods to help remove differences in hospital performance that are due to random chance. Because unmeasured differences in patient mix and other factors besides quality of care may influence hospital mortality, these measures can benefit significantly from use in conjunction with other sources of data on hospital quality. For example, medical chart reviews and other types of electronic clinical data collection (e.g., laboratory test results) can be used to better adjust for severity and comorbidity in comparisons across hospitals. Record reviews may also be helpful for identifying weaknesses in processes of care that are correlated with mortality. Our empirical analysis also showed that many of the mortality indicators are significantly related to each other, suggesting that information on more general aspects of hospital quality (e.g., staffing ratios, procedures to avoid medication errors)

may be useful to examine in hospitals with unusual performance. Better information on post-hospitalization morbidity can be obtained by linking hospital records longitudinally or by surveying patients, and better information on post-admission mortality can be obtained by linking death certificate data. Finally, analyses of hospital outpatient data (particularly ambulatory surgery and emergency room data) in conjunction with inpatient discharge data can help to determine whether the mortality measures reflect differences in outpatient practices.

### **Provider level Utilization Indicators**

The hospital utilization indicators not only show large variations across hospitals; they also show some relationships to other hospital quality indicators and thus may be helpful as “proxies” for other aspects of care. As with the HCUP mortality indicators, these indicators are generally likely to be most useful as a “screen” for further evaluations using supplemental data to determine whether utilization is truly inappropriate. However, these indicators are generally more precisely measured, so that “smoothing” methods are less critical for identifying systematic differences in hospital performance. Additional data collection (e.g., chart review) is also less critical for some of these measures. For example, incidental appendectomy is almost always inappropriate and bilateral catheterization is usually inappropriate, though review of some of the cases performed might identify valid exceptions. For the other utilization indicators, detailed clinical guidelines on appropriate use have been developed and could be applied to determine whether hospitals that appear to have high rates are in fact treating an unusually large number of inappropriate or questionable cases.

### **Area Level Utilization Indicators**

The area utilization indicators all demonstrate substantial differences in procedure rates across MSAs that are apparent even without sophisticated statistical methods. For all of these indicators, detailed clinical guidelines exist for judging the appropriateness of procedure use. Such guidelines can be applied to sample cases from hospitals that make large contributions to high area rates, to help identify specific opportunities for safely lowering rates. For some of the area utilization indicators, e.g. CABG rate, previous studies have shown little variation in inappropriate procedure use and significant underutilization in “necessary” cases, so any effort to lower procedure rates should be undertaken very cautiously. However, in conjunction with the other recommended CABG indicators, this indicator can help provide a relatively comprehensive picture of CABG utilization and outcomes in an area and so may be helpful for public health purposes. Further investigation of area rate differences might also involve collecting information on patient residence, to identify and exclude patients from outside the area from the area rate calculations. Patient residence information could also be used to provide a “proxy” (based on zip code) for patient income and other characteristics of the area that may influence rates.

### **Area Level Avoidable Hospitalizations/ ACSC**

All of the recommended ACSC indicators also show considerable variation across areas, though for some of the indicators, smoothing methods should be used to avoid erroneous classification of outliers. Unfortunately, for many of the ACSC indicators, the available literature on causes of area rate differences is limited. Nonetheless, some further investigations are likely to provide useful insights. The vast majority of patients hospitalized with a subset of the ACSCs are elderly (e.g., dehydration, pneumonia). For these conditions, complementary analyses of data from the Medicare program, which include longitudinal records of both inpatient and outpatient care, can

provide further insights about whether high area rates are associated with less use of outpatient care. Even though HCUP data lack detail, they are much more complete in terms of providing information on Medicare beneficiaries enrolled in managed care plans (historically, managed care plans in Medicare have not reported inpatient or outpatient encounter data). Thus, Medicare and HCUP data may be complementary, especially in areas with high rates of managed care enrollment among the elderly. As with the area utilization indicators, additional information on patient residence can support analyses of the impact of “leakage” in and out of MSAs, and analyses of effects of socioeconomic and other area characteristics on rates. In addition, information abstracted from medical records can provide evidence on whether some of the admissions might have been avoidable, and on whether hospitals and areas differ in their ability to manage some of the ACSCs effectively on an outpatient basis.

### **Summary**

Extensive literature review and empirical evaluation identified 45 quality indicators, out of over 200 indicators inventoried, that can be used with hospital administrative data, similar to HCUP data. These 45 indicators had the best face validity and empirical performance of all evaluated indicators. The results of that evaluation are presented in this report. In addition, the indicators are available in a software package, written in SAS programming language. These quality indicators are intended as quality screens or tools to identify potential problem areas in health care quality, primarily providing an impetus for further investigation. The report discusses the proper use of these indicators, making indicator specific recommendations for further investigations. Such recommendations include analyzing indicators in context of related indicators, using additional data or chart review to identify quality problems, and further investigating sources of potential bias. For reasons fully described in the report, these indicators may not be appropriate for public accountability programs, at least without further attention to the potential limitations and sources of bias. We conclude by setting forth suggestions for future enhancements to HCUP data and recommendations for future research on quality indicators.