*Technical Review 4*

# Refinement of the HCUP Quality Indicators

Agency for Healthcare Research and Quality

*www.ahrq.gov*
***Technical Review***
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

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

ISBN 1-58763-043-5

## Structured Abstract

**Objectives.** In 1994, the Agency for Healthcare Research and Quality (AHRQ) developed the Healthcare Cost and Utilization Project (HCUP) Quality Indicators (QIs), in response to the increasing demand for information regarding the quality of health care. These measures, based on discharge data, were intended to flag potential quality problems in hospitals or regions. The purpose of this project is to refine the original set of HCUP QIs (HCUP I) and recommend a revised indicator set (HCUP II). Specifically this project aims to 1) identify quality indicators reported in the literature and in use by health care organizations, 2) evaluate both HCUP I QIs and other indicators using literature review and novel empirical methods, and 3) make recommendations for the HCUP II QI set and further research. The project deferred evaluation of indicators of complications to a separate study and report.

**Evaluation framework.** Potential and current QIs were evaluated according to six criteria.

(1) Face validity. An adequate QI must have sound clinical and or empirical rationale for its use, and measure an important aspect of quality that is subject to provider or health care system control.
(2) Precision. An adequate QI should have relatively large variation among providers that is not due to random variation or patient characteristics.
(3) Minimum bias. The indicator should not be affected by systematic differences in patient case-mix. In instances where such systematic differences exist, an adequate risk adjustment system should be available based on HCUP discharge data.
(4) 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.
(5) Fosters real quality improvement. The indicator should not create incentives or rewards for providers to improve measured performance without truly improving quality of care.
(6) 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.

**Literature review.** Two separate literature reviews were performed using MEDLINE. The first search (Phase 1) utilized a structured methodology, designed to locate quality indicators developed since 1994 and reported in the literature. The search terms used were “hospital, statistic and methods” and “quality indicator.” Indicators were also identified through web searches and contacts with quality measurement experts.A second search (Phase 2) was used to evaluate each indicator according to the evaluation framework above. MEDLINE (1990-2001) was searched for relevant articles discussing one of the six evaluation framework criterion for selected QIs.

**Empirical evaluation.** Selected indicators were tested using a series of empirical analyses designed to test precision (signal variance, provider- or area-level variance,

signal-to-noise ratio, and R-square), minimum bias (impact of risk adjustment measured by Spearman's rank correlation, percentage remaining in extreme deciles, absolute change in performance, and percent changing more than two deciles), and construct validity (Pearson correlation and factor analysis). Each indicator was assigned a summary score for empirical performance using results from the precision, and to a lesser extent bias tests.

**Selection criteria.** Due to resource constraints, only a portion of the over 200 identified indicators were evaluated comprehensively (all empirical analyses tests and detailed literature review). Indicators were selected for comprehensive evaluation based on the following criteria:

(1) The indicator must have adequate clinical rationale.
(2) The measured event must be somewhat frequent and occur in an adequate number of providers or areas.
(3) The indicator must perform adequately well on preliminary tests of precision.

**Main results.** Forty-five indicators were recommended for use in the HCUP II QI set, including volume, mortality, utilization and ambulatory care sensitive condition measures. Each indicator is appropriate for use as a "quality screen," meaning as an initial tool to identify potential quality problems. These indicators would not be expected to definitively distinguish low quality providers or areas from high quality providers or areas. The empirical performance of each indicator was evaluated; summary empirical scores ranged from 3 to 23 out of a possible 26. All indicators are recommended with specific caveats of use, identified primarily through literature review. Most volume and utilization indicators are best used as proxy measures of quality. Some indicators carry substantial selection bias due to the elective nature of some admissions and procedures. Other indicators are subject to information bias, due to the inability to track post-hospitalization mortality rates. Confounding bias, due to systematic differences in case mix, was found to be a concern for some indicators. Further, many indicators have limited evidence supporting their construct validity; others are somewhat imprecise and require smoothing techniques. Finally, some indicators may create perverse incentives for over- or under-utilization.. Specifics of the caveats of use can be found in the Executive Summary of this report. Ten indicators are recommended for use only in conjunction with other indicators.

Twenty-five of the indicators are provider level indicators, meaning that they evaluate quality of care at the provider (in this case, hospital) level. These indicators include seven procedure volume indicators (AAA repair, carotid endarterectomy, CABG, esophageal resection, pancreatic resection, pediatric heart surgery, and PTCA), five procedure utilization indicators (Cesarean section rate, incidental appendectomy rate, bi-lateral heart catheterization rate, VBAC rate, and laparoscopic cholecystectomy rate), six in-hospital medical mortality indicators (AMI, CHF, GI hemorrhage, hip fracture, pneumonia and stroke), and seven in-hospital provider mortality indicators (AAA repair, CABG, craniotomy, esophageal resection, hip replacement, pancreatic resection, and pediatric heart surgery).

Twenty of the recommended indicators are area-level indicators, meaning that they have population denominators and likely measure quality of the health care system in an area. These indicators include four procedure utilization indicators (CABG, hysterectomy, laminectomy, and PTCA), and sixteen ambulatory care sensitive condition indicators (dehydration, bacterial pneumonia, urinary tract infection, perforated appendix, angina, asthma, COPD, CHF, diabetes short term complications, uncontrolled diabetes, diabetes long term complications, lower extremity amputation in diabetics, hypertension, low birth weight, pediatric asthma and pediatric gastroenteritis).

**Conclusions and future research.** This project identified 45 indicators that are promising for use as quality screens, demonstrating through literature review and empirical analyses that useful information regarding quality of health care can be gleaned from routinely collected administrative data. However, these indicators have important limitations and could benefit from further research. Techniques such as risk adjustment and multivariate smoothing are currently available to reduce the impact of some of these limitations, but other limitations remain.

There are two major recommendations for further action and research – (1) the improvement of HCUP data and subsequently the HCUP QIs to address some of the noted limitations, and (2) further research into quality measurement and the reality of these limitations. The HCUP QIs could benefit from the inclusion of additional data, some of which is now routinely available in some states. Important additions to data include hospital outpatient; emergency room and ambulatory surgery data; linkages to vital statistics such as death records to track post-hospitalization deaths for mortality indicators or birth records for better obstetric risk adjustment; and additional clinical data to improve the risk adjustment available. In addition, research into quality measurement should continue. The relationships underlying the validity of volume measures and utilization measures needs to be revisited periodically to assure validity. Further, research surrounding the construct validity of indicators is essential. Finally, further research is needed regarding risk adjustment of indicators, and how alternative risk adjustment methods affect indicators.

---

This document is in the public domain and may be used and reprinted without permission.

Suggested citation:

Davies SM, Geppert J, McClellan M, et al. Refinement of the HCUP Quality Indicators. Technical Review Number 4 (Prepared by UCSF-Stanford Evidence-based Practice Center under Contract No. 290-97-0013). AHRQ Publication No. 01-0035. Rockville, MD: Agency for Healthcare Research and Quality. May 2001.

**CONTENTS**

**EXECUTIVE SUMMARY** ..... **3**

**CHAPTER 1: INTRODUCTION** ..... **25**

1.A. HEALTHCARE COST AND UTILIZATION PROJECT (HCUP) QIS ..... 26
1.B. EVIDENCE PROJECT TO IMPROVE HCUP QIS ..... 27
1.C. ANTICIPATED USES OF THIS EVIDENCE REPORT ..... 28

**CHAPTER 2: METHODS** ..... **30**

2.A. SEMI-STRUCTURED INTERVIEWS ..... 37
2.B. QUALITY INDICATORS EVALUATION FRAMEWORK ..... 37
2.C. LITERATURE REVIEW OF QUALITY INDICATORS ..... 46
*Phase 1: Identification of Indicators* ..... *46*
*Phase 2: Evaluation of Indicators* ..... *49*
2.D. RISK ADJUSTMENT OF HCUP QUALITY INDICATORS ..... 56
2.E. EMPIRICAL METHODS ..... 63

**CHAPTER 3: RESULTS** ..... **73**

3. A. SUMMARY OF EVIDENCE FOR INDICATORS ..... 74
*3.A.1. Initial Empirical Evaluation* ..... *74*
*3.A.2. Summary of Results* ..... *76*
*3.A.3. Summary Tables of Evidence by Indicator* ..... *94*
*3.A.4. Summary Tables of Evidence by Empirical Tests* ..... *147*
3.B. OVERVIEW OF SEMI-STRUCTURED INTERVIEWS ..... 166
3.C. REVIEW OF RISK ADJUSTMENT APPROACH ..... 171
3.D. GENERAL EVIDENCE FROM THE LITERATURE BY INDICATOR TYPE ..... 178
3.E. DETAILED EVIDENCE BY INDICATOR ..... 185
*3.E.1 Volume Measures* ..... *185*
Indicator 1: Abdominal Aortic Aneurysm (AAA) Repair Volume ..... 185
Indicator 2: Carotid Endarterectomy (CE) Volume ..... 190
Indicator 3: Coronary Artery Bypass Graft (CABG) Volume ..... 194
Indicator 4: Esophageal Resection Volume ..... 198
Indicator 5: Pancreatic Resection Volume ..... 202
Indicator 6: Pediatric Heart Surgery Volume ..... 206
Indicator 7: Percutaneous Transluminal Coronary Angioplasty (PTCA) Volume ..... 210
*3.E.2 Provider-Level Utilization Measures* ..... *215*
Indicator 8: Cesarean Section Delivery Rate ..... 215
Indicator 9: Incidental Appendectomy Among the Elderly Rate ..... 220
Indicator 10: Bi-lateral Cardiac Catheterization Rate ..... 224
Indicator 11: Successful Vaginal Birth After Cesarean Section (VBAC) Rate ..... 228
Indicator 12: Laparoscopic Cholecystectomy Rate ..... 233
*3.E.3. Area-level utilization measures* ..... *238*
Indicator 13: Coronary Artery Bypass Graft (CABG) Rate ..... 238
Indicator 14: Hysterectomy Rate ..... 242
Indicator 15: Laminectomy and/or Spinal Fusion Rate ..... 246
Indicator 16: Percutaneous Transluminal Coronary Angioplasty (PTCA) Rate ..... 250
*3.E.4. Ambulatory Care Sensitive Condition Measures* ..... *254*
Indicator 17: ACSC: Dehydration Admission Rate ..... 258
Indicator 18: ACSC: Bacterial Pneumonia Admission Rate ..... 261
Indicator 19: ACSC: Urinary Infection Admission Rate ..... 265
Indicator 20: ACSC: Perforated Appendix Admission Rate ..... 268
Indicator 21: ACSC: Angina Without Procedure Admission Rate ..... 272
Indicator 22: ACSC: Adult Asthma Admission Rate ..... 275
Indicator 23: ACSC: Chronic Obstructive Pulmonary Disease (COPD) Admission Rate ..... 281
Indicator 24: ACSC: Congestive Heart Failure (CHF) Admission Rate ..... 286
Indicator 25: ACSC: Diabetes – Short Term Complications Admission Rate ..... 290

Indicator 26: ACSC: Uncontrolled Diabetes Admission Rate ..... 295
Indicator 27: ACSC: Diabetes – Long Term Complications Admission Rate ..... 297
Indicator 28: ACSC: Hypertension Admission Rate ..... 301
Indicator 29: ACSC: Lower Extremity Amputation Rate ..... 304
Indicator 30: ACSC: Low Birth Weight Rate ..... 309
Indicator 31: ACSC: Pediatric Asthma Admission Rate ..... 313
Indicator 32: ACSC: Pediatric Gastroenteritis Admission Rate ..... 318
*3.E.5. In-Hospital Mortality Measures ..... 322*
Indicator 33: Acute Myocardial Infarction (AMI) Mortality Rate ..... 322
Indicator 34: Congestive Heart Failure (CHF) Mortality Rate ..... 328
Indicator 35: Gastrointestinal (GI) Hemorrhage Mortality Rate ..... 332
Indicator 36: Hip Fracture Mortality Rate ..... 336
Indicator 37: Pneumonia Mortality Rate ..... 341
Indicator 38: Acute Stroke Mortality Rate ..... 346
*3.E.6. Post-Procedural Mortality Measures ..... 351*
Indicator 39: Abdominal Aortic Aneurysm (AAA) Repair Mortality Rate ..... 351
Indicator 40: Coronary Artery Bypass Graft (CABG) Mortality Rate ..... 355
Indicator 41: Craniotomy Mortality Rate ..... 360
Indicator 42: Esophageal Resection Mortality Rate ..... 363
Indicator 43: Hip Replacement Mortality Rate ..... 366
Indicator 44: Pancreatic Resection Mortality Rate ..... 370
Indicator 45: Pediatric Heart Surgery Mortality Rate ..... 373

**CHAPTER 4: CONCLUSIONS AND FUTURE RESEARCH ..... 377**

**REFERENCES ..... 382**

## APPENDICES

**PEER REVIEWERS ..... APPENDIX 1**

**CONTACT LIST FOR PHONE INTERVIEWS ..... APPENDIX 2**

**LIST OF ABSTRACTED ARTICLES ..... APPENDIX 3**

**DATA COLLECTION INSTRUMENTS ..... APPENDIX 4**

**INVENTORY OF INDICATORS ..... APPENDIX 5**

**INDICATOR DEFINITIONS ..... APPENDIX 6**

**EMPIRICAL EVALUATION RESULTS ..... APPENDIX 7**

**LITERATURE TABLES FOR UTILIZATION AND ACSC INDICATORS ..... APPENDIX 8**

**LIST OF ACRONYMS ..... APPENDIX 9**

# LIST OF FIGURES AND TABLES

**Executive summary**

Figure 1S. Flow chart of the identification of recommended indicators 10

Table 1S. Provider indicator list 15
Table 2S. Area indicator list 18

**Body of report**

Figure 1. Abbreviated diagram of literature review 31
Figure 2. Example search strings 50
Figure 3. Example search 51

Flow Diagram 1. Literature review strategy 54

Table 1. Precision Tests 70
Table 2. Bias Tests 71
Table 3. Construct Validity Tests 72
Table 4. List of HCUP I indicators, and inclusion status in HCUP II. 75
Table 5. List of other indicators, tested but not recommended due to low precision 75
Table 6. Provider indicator list 77
Table 7. Area indicator list 79
Table 8. Example clinical groupings 93
Table 9. Summary evidence table for volume indicators 95
Table 10. Summary evidence table for provider level utilization indicators 101
Table 11. Summary evidence table for area-level utilization indicators. 106
Table 12. Summary evidence table for area-level ACSC indicators. 111
Table 13. Summary evidence table for mortality indicators 129
Table 14. Thresholds for indicators 147
Table 15. Volume distribution by year 147
Table 16. Percentage of procedures and volume 149
Table 17. Year to year persistence of high volume status 150
Table 18. Correlations between volume and mortality indicators 151
Table 19. Precision – Provider Indicators 152
Table 20. Precision – Area Indicators 153
Table 21. Minimum Bias – Provider Indicators 157
Table 22. Minimum Bias – Area Indicators 158
Table 23. Summary of Factor Analysis Results, Provider-level indicators 163
Table 24. Summary of Factor Analysis Results, Area-level ACSC indicators 165
Table 25. Summary of Factor Analysis Results, Area-level utilization indicators 165
Table 26. Summary of Telephone Contacts August 1999 through March 2001 168
Table 27. Risk adjustment statistics from the literature 175

# EVIDENCE REPORT

**Comprehensive technical report of the project**

# 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.

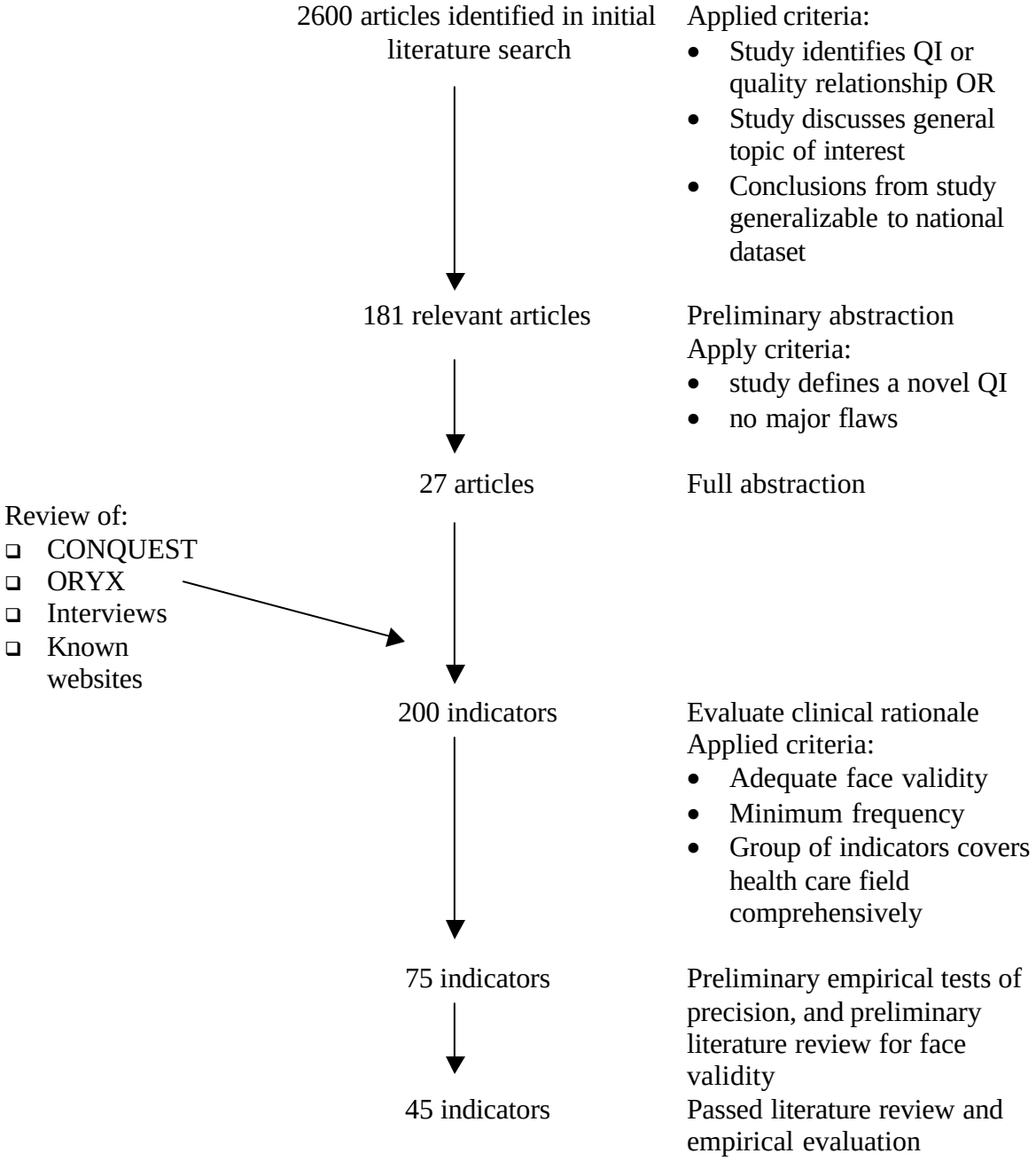
## Results

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 in 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**

| Provider Level Indicator Name | Type of Indicator | Indicator Description | Empirical Rating[a] | Risk Adjustment Recommended | Caveats of use from the literature review[b] | Better quality may be associated with: |
|---|---|---|---|---|---|---|
| AAA repair volume (#1) | Volume | Raw volume compared to annual thresholds (20, 32 procedures) | Not applicable | Not applicable | ŏ Proxy<br>? Easily manipulated | Higher rates |
| Carotid endarterectomy volume (#2) | Volume | Raw volume compared to annual thresholds (50, 101 procedures) | Not applicable | Not applicable | ŏ Proxy<br>ŏ Easily manipulated | Higher rates |
| CABG volume (#3) | Volume | Raw volume compared to annual thresholds (100, 200 procedures) | Not applicable | Not applicable | ŏ Proxy<br>? Easily manipulated | Higher rates |
| Esophageal resection volume (#4) | Volume | Raw volume compared to annual thresholds (6, 7 procedures) | Not applicable | Not applicable | ŏ Proxy<br>? Easily manipulated | Higher rates |
| Pancreatic resection volume (#5) | Volume | Raw volume compared to annual thresholds (10, 11 procedures) | Not applicable | Not applicable | ŏ Proxy<br>? Easily manipulated | Higher rates |
| Pediatric heart surgery volume (#6) | Volume | Raw volume compared to annual thresholds (100 procedures) | Not applicable | Not applicable | ŏ Proxy<br>? Easily manipulated | Higher rates |
| PTCA volume (#7) | Volume | Raw volume compared to annual thresholds (200, 400 procedures) | Not applicable | Not applicable | ŏ Proxy<br>? Selection bias<br>ŏ 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<br>? Unclear construct validity<br>? Unclear benchmark | Lower rates[c] |
| Incidental appendectomy among elderly rate (#9) | Utilization | Number of incidental appendectomies per 100 abdominal surgeries | 13<br>Smoothing recommended | Age and sex | ? Unclear construct validity<br>? Easily manipulated | Lower rates |

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

| Provider Level Indicator Name | Type of Indicator | Indicator Description | Empirical Rating[a] | Risk Adjustment Recommended | Caveats of use from the literature review[b] | Better quality may be associated with: |
|---|---|---|---|---|---|---|
| (#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<br>ŏ Confounding bias<br>? Unclear construct validity<br>? Easily manipulated | Lower rates |
| Craniotomy mortality (#41) | Post-procedural Mortality | Number of deaths per 100 craniotomies | 6 Smoothing recommended | APR-DRG | ŏ Confounding bias<br>? 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<br>? 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<br>? Confounding bias<br>? 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<br>? Unclear construct validity | Lower rates |
| Pediatric heart surgery mortality (#45)[d] | Post-procedural Mortality | Number of deaths per 100 heart surgeries in patients under age 18 years | 3 Smoothing recommended | APR-DRG | ŏ Confounding bias<br>? Unclear construct validity<br>Unclear benchmark | Lower rates |

[a]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.
[b]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.
[c]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.
[d]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[a] | Risk Adjustment Recommended | Literature review findings[b] | Better quality may be associated with: |
|---|---|---|---|---|---|---|
| CABG rate (#13)[d] | Utilization | Number of CABGs per 100,000 population | 19 | Age and sex | ☹ Proxy<br>☹ Unclear construct validity<br>☹ Unclear benchmark | Lower rates[c] |
| Hysterectomy rate (#14) | Utilization | Number of hysterectomies per 100,000 population | 22 | Age and additional factors such as parity. | ☹ Proxy<br>? Confounding bias<br>☹ Unclear construct validity<br>☹ Unclear benchmark | Lower rates[c] |
| Laminectomy rate (#15) | Utilization | Number of laminectomies per 100,000 population | 20 | Age and sex | ☹ Proxy<br>☹ Unclear construct validity<br>☹ Unclear benchmark | Lower rates[c] |
| PTCA rate (#16)[d] | Utilization | Number of PTCAs per 100,000 population | 19 | Age and sex | ☹ Proxy<br>? Selection bias<br>☹ Unclear construct validity<br>☹ Unclear benchmark | Lower rates[c] |
| Dehydration admission rate (#17)[d] | ACSC | Number of admissions for dehydration per 100,000 population | 14 | Age and sex | ? Proxy<br>? Unclear construct validity<br>? Easily manipulated<br>☹ Unclear benchmark | Lower rates[c] |
| Bacterial pneumonia admission rate (#18)[d] | ACSC | Number of admissions for bacterial pneumonia per 100,000 population | 17 | Age and sex | ? Proxy<br>? Unclear construct validity<br>? Easily manipulated<br>☹ Unclear benchmark | Lower rates[c] |
| Urinary infection admission rate (#19)[d] | ACSC | Number of admissions for urinary infection per 100,000 population | 11 | Age and sex | ? Proxy<br>? Unclear construct validity<br>? Easily manipulated<br>☹ Unclear benchmark | Lower rates[c] |
| Perforated appendix admission rate (#20) | ACSC | Number of admissions for perforated as a share of all admissions for appendicitis within an area | 17<br>Smoothing recommended | Age and sex | ? Proxy | Lower rates |
| Angina admission | ACSC | Number of admissions for | 19 | Age and sex | ? Proxy | Lower rates[c] |

| Area Level Indicator Name | Type of Indicator | Indicator Description | Empirical Rating[a] | Risk Adjustment Recommended | Literature review findings[b] | Better quality may be associated with: |
|---|---|---|---|---|---|---|
| rate (#21)[d] | | angina per 100,000 population | | | ? Unclear construct validity<br>? Easily manipulated<br>ŏ Unclear benchmark | |
| Adult asthma admission rate (#22) | ACSC | Number of admissions for asthma in adults per 100,000 population | 16 | Age and sex | ? Proxy<br>? Easily manipulated<br>ŏ Unclear benchmark | Lower rates[c] |
| 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<br>? Confounding bias<br>? Easily manipulated<br>ŏ Unclear benchmark | Lower rates[c] |
| CHF admission rate (#24) | ACSC | Number of admissions for CHF per 100,000 population | 14 | Age and sex. | ? Proxy<br>? Easily manipulated<br>ŏ Unclear benchmark | Lower rates[c] |
| 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<br>? Confounding bias | Lower rates |
| Uncontrolled diabetes admission rate (#26)[d] | ACSC | Number of admissions for uncontrolled diabetes per 100,000 population | 14 | Age and sex. Potential supplement includes population diabetes incidence rates. | ? Proxy<br>? Confounding bias<br>? 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<br>? Confounding bias<br>? Easily manipulated<br>ŏ Unclear benchmark | Lower rates |
| Hypertension admission rate (#28) | ACSC | Number of admissions for hypertension per 100,000 population | 14 | Age and sex | ? Proxy<br>? Easily manipulated<br>ŏ Unclear benchmark | Lower rates[c] |
| 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[a] | Risk Adjustment Recommended | Literature review findings[b] | Better quality may be associated with: |
|---|---|---|---|---|---|---|
| amputation in diabetics rate (#29)[d] | | 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)[d] | ACSC | Number low birth weight births as a share of all births in an area | 11 out of 16[e] Smoothing recommended | None available in HCUP. Potential supplement include clinical, link to mother's record, or link to birth record. | ? Proxy<br>? Confounding bias<br>ŏ 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<br>? Easily manipulated<br>ŏ Unclear benchmark | Lower rates[c] |
| Pediatric gastroenteritis admission rate (#32)[d] | ACSC | Number of admissions for pediatric gastroenteritis per 100,000 population | 17 | Age and sex | ? Proxy<br>? Unclear construct validity<br>? Easily manipulated<br>ŏ Unclear benchmark | Lower rates[c] |

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

# Chapter 1: Introduction

This Evidence Report is organized into four major chapters – Introduction, Methods, Results, and Conclusions. This evidence report is the first of two reports. This report evaluates indicators relating to hospital structure, processes (though few process indicators can be derived from administrative data), and two specific outcomes – hospitalization for ambulatory care sensitive conditions and mortality. The second report will evaluate complications indicators, with a special focus on patient safety.

The first section of this introduction chapter provides background on a quality measurement tool, the Healthcare Cost and Utilization Project Quality Indicators (QIs), developed in 1994 at the Agency for Health Care Policy and Research (AHCPR, now the Agency for Healthcare Research and Quality, AHRQ). The original 33 indicators, hereafter referred to as the HCUP I QIs, were based on inpatient hospitalization data, and designed to highlight potential quality concerns and to target areas for further analysis. Current users of these indicators include hospital associations, state health departments, data organizations, and individual hospitals. Because there have been new developments in the quality measurement field and because of requests from users for enhancements to the HCUP I QIs, AHRQ decided to request proposals for the improvement of the current tool. In response to this request, the UCSF-Stanford Evidence-based Practice Center undertook and reports here an extensive project to determine the evidence base for a refinement of the HCUP I QIs. The research team also developed software for the HCUP II QIs to implement the recommended indicators described in this report.

The second section of this chapter introduces the general approach to improving the HCUP I QIs. The project team searched the literature and contacted experts in the field to identify potential new indicators, developed evaluation criteria, followed by evaluation of current and potential new indicators using both literature sources and analyses of the HCUP data sets. The resulting quality indicators recommended in this report are at minimum "screening" tools that should flag potential quality problems. Each indicator has caveats of use, and these should be taken into account regardless of the application of these indicators. As these measures were evaluated as "screening tools," they have not been evaluated as definitive measures of provider quality. They most likely cannot definitively distinguish poor quality providers from high quality providers, but rather may illuminate potential quality problems for consideration of further investigation.

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."[1] We further focused on the clinical domains of potential underuse, overuse, and misuse,[2] and excluded potential indicators based on patient satisfaction, health professional satisfaction, or cost containment.

The report is expected to be of value to those interested in using the HCUP II QIs since it documents the evidence project to develop recommendations for improvements to the HCUP I quality indicators. The report excludes a review of indicators relating to potential complications of care (including 11 HCUP I indicators) because this set will be included in a separate evidence report covering complications and patient safety

indicators. Further details about likely use of and audiences for this report can be found in the last section of the introduction chapter.

## 1.A. Healthcare Cost and Utilization Project (HCUP) QIs

The demand for information on quality in health care has risen sharply over the past decade. In response to this demand, the Healthcare Cost and Utilization Project (HCUP I) Quality Indicators (QIs) were developed at the Agency for Healthcare Research and Quality (AHRQ, formerly Agency for Health Care Research and Policy) in 1994. HCUP is a federal-state-private collaboration to build uniform databases from administrative hospital data collected by state data organizations and hospital associations. These data are gathered by AHRQ, converted into a uniform format, and released for research purposes. Currently, 22 states participate in HCUP. In the early 1990s, the HCUP state partners requested assistance in making greater use of the administrative data they collected. In response, AHRQ researchers conducted a literature review to identify measures of quality that could be constructed using only information routinely found on hospital discharge abstracts and claims. The HCUP I Quality Indicator software was developed and disseminated.

The 1994 HCUP indicator set (HCUP I), which is in use currently, consists of 33 indicators constructed using administrative data available in the Nationwide Inpatient Sample (NIS), one of the HCUP databases. Included in the set are indicators of procedure utilization, ambulatory care sensitive condition admissions, post-operative and other complications of care, and mortality. Developers of this indicator set compiled a list of quality indicators used by a variety of organizations, and then selected indicators based on simplicity, feasibility, evidence from a review of the literature, and coverage of a range of potentially affected patient populations. The 33 resulting quality indicators were designed to highlight quality concerns, particularly in clinical performance, and to target areas for further analysis. As a result, the indicators were not intended as definitive measures of quality problems, but rather as screens for use in quality improvement. As screening tools, these indicators would serve as a first-round flag of potential quality problems, which should be investigated further by other methods, such as chart review.

Since indicators with complex multivariate adjustments can be difficult to understand and to interpret, the original HCUP I indicators were developed without such techniques, and with the objective of simplicity. Two approaches were utilized to control for differences in severity of illness or case mix across hospitals. First, numerators and denominators defined to assess rates for each indicator were restricted to homogenous groups whenever feasible. For example, QIs examining in-hospital mortality following common elective procedures restrict the populations at risk to uncomplicated cases only. The second approach to controlling for differences in case mix employed standardization based on the Major Diagnostic Code (MDC) mix of the hospital compared to the state's MDC mix.[3]

HCUP I QIs were released publicly in 1994 as a set of statistical programs.[4] AHRQ has disseminated widely this software tool in both SAS and SPSS statistical package formats. Current users of these indicators include hospital associations, state health departments, data organizations, and individual hospitals. In general, these indicators are used for preliminary quality screening, internal reporting, and, in a few cases, non-public benchmarking. Specific examples of use include:

(1) hospitals have refined their patient selection criteria for procedures to improve appropriate usage;
(2) hospitals have reviewed medical records to verify the presence of poor outcomes and to investigate the reasons behind the event identified by the QIs;
(3) the QIs have served as a springboard for hospitals to identify community health care needs and plan for particular services (e.g., one hospital worked with the surrounding community to strengthen outpatient services for diabetes patients after seeing QI data);
(4) a project with one state chapter of the American College of Obstetricians Gynecologists was initiated to reduce adverse outcomes associated with low birth weight infants; and
(5) use of the QIs has provided hospitals the opportunity to verify and improve the accuracy of the data they reported.

Although the original 33 HCUP I QIs were designed using a multi-state hospital discharge data set, they can provide information on care both within and outside hospitals. Some of the QIs offer hospitals self-assessment opportunities in the areas of potentially avoidable adverse hospital outcomes and of potentially inappropriate utilization of hospital procedures. Other QIs cover potentially avoidable hospital admissions, and therefore provide public health officials insight into the quality and appropriateness of medical care in the community.

## 1.B. Evidence Project to Improve HCUP QIs

Since the time of the original HCUP I development work in 1994, numerous managed care organizations, state Medicaid agencies and hospital associations, quality improvement organizations, the National Committee on Quality Assurance (NCQA), the Joint Commission on the Accreditation of Healthcare Organizations (JCAHO), academic researchers and others have contributed substantially to the knowledge base of hospital quality indicators. Based on input from current users and on 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.

Three primary goals were established to accomplish this task: 1) Identify potential new indicators, 2) Evaluate existing HCUP I indicators and potential indicators using both literature review and empirical analyses of indicator performance, and 3) Examine the need for risk adjustment of recommended indicators. As part of the indicator identification goal, AHRQ requested that additional QIs cover populations and conditions that are not well represented in the HCUP I QIs. In particular, the EPC was expected to review the evidence for indicators in four priority expansion areas: 1) Pediatric conditions, 2) Medical conditions including chronic illnesses, 3) Potentially avoidable hospitalizations, particularly in Healthy People 2010 objective areas,[5] and 4) New procedures not covered in the HCUP I QIs. For the indicator evaluation goal, AHRQ requested that the current QIs be reviewed for their appropriateness in terms of choice of

denominator (discharge-based versus population-based) and for their overall validity as a measure of quality. For the final goal, the EPC was asked to evaluate more sophisticated methods of risk adjustment using hospital discharge data, based on input from experts in the field, and literature reviews.

The team designed a series of investigations to accomplish these goals. These included telephone interviews of those knowledgeable about quality measurement, two phases of extensive literature reviews, and a series of empirical analyses using the HCUP data sets. 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, hereafter referred to as the HCUP II 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?
- Given the potential importance of risk adjustment for appropriate interpretation of the indicators, 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?

## 1.C. Anticipated Uses of this Evidence Report

The approach to identification and evaluation of QIs presented in this report underpins the development of the HCUP II QIs. 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 and risk adjustment strategies, the report may serve as a springboard for commentary on proposed recommendations for specific improvements to the HCUP I QIs.

This report is likely also to serve as a reference for background material on quality measurement, and as a review of the current state of quality indicators and risk adjustment methods. In addition, it documents a novel application of evidence-based methods to quality indicator evaluation, adapting the traditional use of these methods for informing clinical practice choices.

We anticipate that the audience for this report is likely to be quite heterogeneous, including health care providers, organizations, policy makers, and public health officials. A primary audience for this report is the AHRQ along with those using the current HCUP

I QI tool since the report provides details about the methods and evaluation results for the recommended indicators. However, the report has been written for a wider audience likely to be interested in improvements to quality measurement in general, and the HCUP QI set specifically. Indicators applicable to health care providers and those actionable by public health officials were both considered in the systematic review and evaluation of potential new indicators for the HCUP II QI set. Therefore, it is likely that each audience will find differing aspects of the report most useful, and as such, we have included a variety of information.

The HCUP II QIs themselves should provide a means of using a readily available data source, hospital administrative data, to provide initial screening of quality of care. These measures can be used to provide national and regional benchmarks against which individual providers, localities, and states can compare themselves. In addition, with an understanding of the limitations of administrative data and with appropriate precautions, the QIs can form a preliminary basis of a quality improvement program. Because of the limitations associated with administrative data, the HCUP QIs are not designed for purchasing decisions and public reporting on individual hospitals.

Potential stakeholders and quality measurement experts were selected to review the report (see Appendix 1 for a list of reviewers). After receiving reviewer comments, the project team revised the report and sent it to an expanded group for a second round of review (also listed in Appendix 1), followed by further revisions. As a result, this report has undergone extensive peer review and subsequent revision based on reviewer comments.

# Chapter 2: Methods

The methods chapter is organized into five sections: 1) Semi-structured interviews, 2) Evaluation Framework, 3) Literature Review, 4) Risk Adjustment Review, and 5) Empirical Methods. A summary description of each of the sections is included below. The remaining sections of the chapter contain the detailed methods.

## Summary of Methods

### Semi-structured interviews

The project team contacted 37 individuals affiliated with hospital associations, business coalitions, state data groups, federal agencies, and academia. These individuals were identified for their potential expertise in quality measurement. We deliberately included quality indicator (QI) users and potential users from a broad spectrum of organizations in both the public and private sectors; the sample was not intended to be representative of any population. Thirty-one individuals agreed to be interviewed about various topics related to quality measurement. Interviews were tailored to the specific expertise of interviewees (listed in Appendix 2). Two team members collected information during interviews regarding quality indicator use, suggested indicators, and other potential contacts.

### Evaluation framework

The project team identified key areas of performance for quality indicators, gathering information from previous evaluations by McGlynn et al.,[6] and Iezzoni et al.,[7-17] and from the telephone interviews. Six evaluation criteria were identified:

1) *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.

2) *Precision.* An adequate quality indicator should have relatively large variation among providers that is not due to random variation or patient characteristics.

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

4) *Construct validity.* The indicator should be related to other indicators or measures intended to measure the same or related aspects of quality.

5) *Fosters real quality improvement.* The indicator should be robust to possible provider manipulation of the system.

6) *Application.* The indicator should have been used in the past, and/or have high potential for working well with other indicators.

### 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 geographical areas (e.g., metropolitan service areas, counties); indicators should also be applicable to a

majority of providers or areas (e.g. not examining specialized care units). The second phase was designed to evaluate those indicators identified in phase 1 using the criteria in the evaluation framework described above.

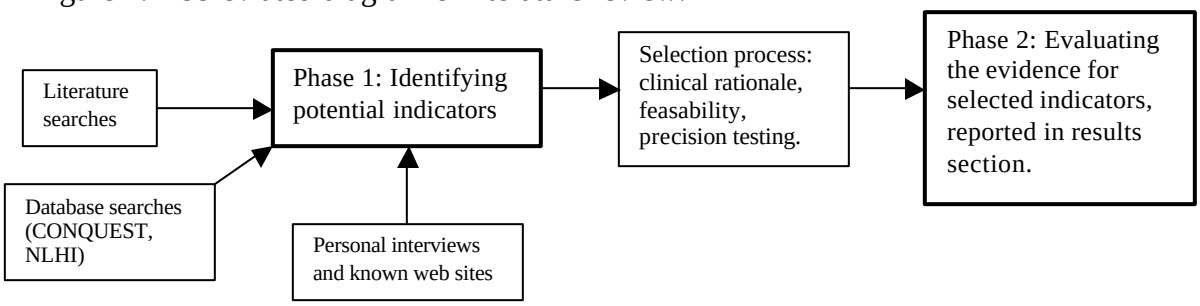
*Phase 1*. To identify potential indicators, we performed a structured review of the literature. Using Medline, we identified the search strategy that returned a test set of known applicable articles in the most concise manner. The final Medical Subject Heading (MeSH) terms used were "hospital, statistic and methods" and "quality indicators." This search resulted in approximately 2600 articles. 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. In the first stage, preliminary abstraction, we evaluated 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 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 (stage 2 of phase 1).

Additional potential indicators were identified using the CONQUEST (COmputerized Needs-oriented QUality Measurement Evaluation SysTem) database, the National Library of Healthcare Indicators (developed by JCAHO), a list of ORYX™ approved indicators provided by the Joint Commission on Accreditation of Healthcare Organizations (JCAHO), and from the telephone interviews.

*Phase 2*. After identifying over 200 potential indicators during Phase 1, we evaluated the indicators. Initially, team members evaluated the clinical rationale of each indicator, and selected the most promising indicators based on certain criteria, including frequency of the event and sound clinical rationale. HCUP I (Healthcare Cost and Utilization Project) indicators were not evaluated in this stage and were automatically selected for the next step of evaluation. Second, those indicators passing the initial screen (including the HCUP I indicators) were subjected to basic empirical tests of precision, as described below. Third, a full literature review was conducted for those indicators with adequate performance on the 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. These indicators also underwent full empirical analyses (see Empirical Methods).

Figure 1. Abbreviated diagram of literature review.




Chapter 2: Methods: Summary

## Risk adjustment review

To identify potential risk adjustment systems, we reviewed the applicable literature regarding risk adjustment systems. We examined risk adjustment approaches for each indicator according to the type of indicator (mortality, utilization, etc.) and the analytic approach, method of development, feasibility of implementation given data availability, and empirical measures of discrimination and calibration. In addition, the semi-structured interviews with potential HCUP users identified several preferences for a risk adjustment system for the HCUP QI set. Users preferred that the system be: 1) open, with published logic, 2) cost-effective, with data collection costs minimized, and additional data collection being well justified, 3) designed using a multiple-use coding system, such as those used for reimbursement, and 4) officially recognized by government, hospital groups or other organizations.

Few risk adjustment systems could be feasibly implemented, given the lack of ambulatory and detailed clinical data. In general, Diagnosis Related Group (DRG) systems tended to fit more of the user preference-based criteria than other alternatives. A majority of users interviewed already used All Patients Refined (APR)-DRGs, and APR-DRGs have been reported to perform well in predicting resource use and death, when compared to other DRG based systems. APR-DRGs also performed as well as or better than other risk adjustment systems for several conditions in a series of studies by Iezzoni et al.[9-13, 17, 18] As a result, we conducted indicator evaluations with the APR-DRG system for two purposes: 1) to determine the impact of measured differences in patient severity on the relative performance of providers, as described in the Empirical Methods' sub-section, and 2) to provide the basis for implementing APR-DRGs as an optional risk adjustment system for hospital level QI measures. The implementation of APR-DRGs is based on an ordinary least squares regression model and is discussed in detail in the full text of the Risk Adjustment Methods sub-section.

Our incorporation of APR-DRGs into the Version 2 software should not be construed as an unequivocal endorsement of this product. Indeed, customized risk-adjustment systems might be more effective than APR-DRGs or any off-the-shelf product. However, it was beyond the scope of this contract to develop customized risk-adjustment systems for each quality indicator.

## 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 (NIS) and the State Inpatient Database (SID) 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 direct 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 full text of the methods section. 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.

For each indicator, the variance can be decomposed into three components: variation within a provider (i.e. differing performance due to differing patient characteristics), variation among providers (i.e. actual differences in performance among providers), and random variation. An ideal indicator would have a substantial amount of the variance explained by between-provider variance, which might possibly result from differences in quality of care, and a minimum amount of random variation.

We conducted four tests of precision (signal standard deviation, provider/area variation share, signal-to-noise ratio and the in-sample r-squared) in order to estimate the magnitude of between-provider variance on each indicator.

(1) Signal Standard Deviation: For the first test of precision, we calculated the signal standard deviation of the QI, which measures the extent to which performance of the QI varies systematically across hospitals or areas. The signal standard deviation is the best estimate of the variation in QI performance that appears to be systematic – truly associated with the hospital or area, and not the result of random variations in patient case mix or environmental conditions. Because many patient characteristics and random events may influence performance of a QI, only some of the apparent variation in performance at the hospital or area level is likely to be systematic. Systematic variation will be larger to the extent that: sample sizes are larger (allowing more precise estimation of the true effect), and patient-associated variation is smaller (allowing more precise estimates for a given sample size). Large systematic variation across hospitals and areas suggests that there are opportunities for further study and quality improvement. By contrast, if the signal standard deviation is small, then there may be little gained by improving the performance of lower-ranked hospitals or areas to the levels achieved by the higher-ranked hospitals or areas.

(2) Provider/Area Share:

Another precision test we calculated is the percentage of signal variance relative to the *total* variance of the QI. In general, the variance in a quality indicator can be divided as follows:

Total Variance = Patient-Level Variance + Provider-Level Variance.

Typically, most variation in the QIs (e.g., whether or not a vaginal birth occurred for a patient with a prior C-section) is associated with patient-level factors, and may have nothing whatsoever to do with providers or areas. In turn, the apparent Provider-Level Variance (or analogously, Area-Level) can be divided into two components:

Provider-Level Variance = Provider-Level Signal Variance + Noise Variance.

That is, some of the variance in a QI rate for a particular provider (adjusted or unadjusted) is a reflection of random chance[1]: in a given year, some hospitals or areas will by chance have more or fewer occurrences of the numerator related to a particular QI. The remaining variance (signal variance), and only the remaining variance, is attributable to systematic or true differences in the QI across providers. The percentage of signal variance relative to total variance will be larger to the extent that: within-hospital patient variance is small; sample sizes are larger; and true differences in performance across hospitals are larger. Unlike the signal standard deviation measure of precision, the percentage of signal variance is a relative measure: relative to the total variation (patient plus provider) in the QI, what fraction appears to be associated with systematic provider differences? Thus, higher percentages suggest enough differences across hospitals or areas that true differences across providers or areas can be distinguished from other "random" influences on QI performance.

(3) Signal-to-Noise Ratio:
The third precision measure we calculated is the Signal-to-Noise ratio in the *provider*- or *area*- level variation in the QI measure. This measure answers the question: of the apparent variation in QIs across providers, what fraction appears to be *truly* related to systematic differences across providers, and not random variations ("noise") from year to year? As noted earlier, the provider-level variation includes both signal variation and random variations in the QI measures across providers ("noise"). In other words, it is the "signal" to "signal plus noise" ratio. As such, the Signal-to-Noise ratio measure is somewhat misnamed. Its definition is:

Signal-to-Noise Ratio = (Provider-Level Signal Variance)/(Provider-Level Total Variance).

In general, if a QI's signal-to-noise ratio is high, then it is likely that apparent variations in performance across providers are not the result of random chance, and careful attention

[1] Random chance as used in this document refers to unsystematic unobservable factors that influence QI performance. Most accurately these factors are neither entirely random nor due to chance, however, statistically they represent the unmeasurable and unsystematic differences that are most likely not due to actual quality differences.

to distinguishing true from random variation across providers will have little impact on the measured performance of a provider. Similarly, for indicators measuring area level rates, the aim is to distinguish between actual and apparent differences between areas.

(4) In-sample R-Squared:
The fourth precision measure identifies the incremental benefit of applying multi-variate signal extraction methods for identifying additional signal on top of the signal-to-noise ratio.

*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. Nonetheless, we recognize that this system is not ideal, because it provides only four severity levels within each base APR-DRG, omits important physiologic and functional predictors, and potentially misadjusts for iatrogenic complications. 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 feasible for the area-level indicators evaluated. We used a range of bias performance measures, most of which have been applied in previous studies.[9-13, 17-19] 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. 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).

(1) Rank Correlation Coefficient: The Spearman correlation coefficient of the rank of the area/hospital without and with risk adjustment gives the overall impact of risk adjustment, on relative provider or area performance.
(2) Average Absolute Value of Change Relative to Mean: The average absolute change in the performance measure of the hospital or area with and without risk adjustment, normalized by the average value of the QI. Thus, it is also a relative measure that theoretically can range from zero (no change) to a much higher value. This measure highlights the amount of absolute change in performance, without reference to other providers' performance. This does not rely on the distribution of performance on this indicator, as relative performance measures do.
(3) Percentage of High That Remains High Decile, or

(4) Percentage of Low Decile That Remains in Low Decile: These two measures report the percentage of hospitals or areas that are in the highest and lowest performance deciles without risk adjustment, that remain there with risk adjustment. A measure that is insensitive to risk adjustment should have rates of 100%, whereas measures for which risk adjustment affects the top- and bottom-performers substantially should have much lower percentages.
(5) Percentage That Change More Than Two Deciles: the percentage of hospitals whose relative rank changes by a substantial distance – more than 20% - with and without risk adjustment. Ideally, risk adjustment would not have a substantial effect for very many hospitals.

*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. Nonetheless, we would expect to see some 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 evaluated to provide some evidence on the nature of the relationship between potential indicators.

# Detailed Methods

## 2.A. Semi-structured Interviews

The project team and previous developers of the Healthcare Cost and Utilization Project Quality Indicators (HCUP I QIs) developed a contact list of individuals associated with hospital associations, business coalitions, state data groups, and federal agencies. This list was designed to include QI users and potential users from a broad spectrum of organizations in both the public and private sectors; it was not intended as a representative sample. All contacts were faxed an introductory letter and asked to participate as advisors on the project with a short telephone interview. This request was well received; only six out of 37 declined participation themselves without suggesting an alternative respondent. Overall, the 31 contacts phoned expressed interest in the study, offering many suggestions and comments (see Appendix 2, List of Contacts). The composition of the 31 interviewees is as follows: three consultants, two federal agency employees, one health plan medical director, five representatives of hospital associations, one international academic researcher, four representatives of private accreditation groups, two representatives of private data groups, two members of professional organizations, five representatives of provider and other private organizations, three representatives of state data groups, and three representatives of other healthcare organizations.

The semi-structured interviews were designed to identify potential indicators, concerns of end users, and other factors important in the development of quality indicators that may not be captured in the published literature. Thus, academic researchers, whose work is more likely to appear in peer reviewed journals were reserved as peer reviewers for the final document. As a result, the results of the semi-structured interviews are not intended to be a non-biased representation of the opinions regarding quality indicators, but rather a sampling of those opinions not likely to be available in the peer-reviewed literature.

Two team members (CV, JG) conducted the interviews. The interviewers solicited information on the development and use of quality indicators by the targeted organizations, as well as other known measures and additional contacts. Interviewers used a semi-structured interview and recorded information from the interview on a data-collection form. Further, some advisors provided the project team with materials regarding quality indicators and the use of HCUP I indicators.

## 2.B. Quality Indicators Evaluation Framework

Six areas were considered essential for evaluating the reliability and validity of a proposed quality indicator. Several sources contributed to the development of the evaluation criteria framework: 1) results of the semi-structured interviews, including the interests and concerns of HCUP users, 2) task order document describing the Agency for Healthcare Research and Quality's (AHRQ) interests, 3) evidence available in the policy and research literature and 4) evidence available through statistical analyses. Our six criteria are quite similar to the criteria for "testing the scientific strength of a measure"

proposed by McGlynn and Asch.[6] They describe a measure as reliable "if, when repeatedly applied to the same population, the same result is obtained a high proportion of the time." They propose evaluating validity in terms of face validity, criterion validity ("an objective assessment of the ability of the measure to predict a score on some other measure that serves as the evaluation criterion"), and construct validity ("whether the correlations between the measure and other measures are of the right magnitude and in the right direction"). We view criterion validity as an assessment of bias (criterion #3), where the "gold standard" measure is purged of bias due to severity of illness. Face validity captures a variety of concepts discussed by McGlynn and Siu[20], including the importance of the condition, the efficacy of available treatments (e.g., the ability of providers to improve outcomes), and the potential for improvement in quality of care.

We assembled the evidence supporting the use of current and candidate HCUP II quality indicators in terms of the following six areas.

| ***Framework for evaluating the quality indicators*** |
|---|
| 1. ***Face validity: does the indicator capture an aspect of quality that is widely regarded as important and subject to provider or public health system control?***<br>2. ***Precision: is there a substantial amount of provider or community level variation that is not attributable to random variation?***<br>3. ***Minimum Bias: is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?***<br>4. ***Construct validity: does the indicator perform well in identifying true (or actual) quality of care problems?***<br>5. ***Fosters real quality improvement: Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?***<br>6. ***Application: Has the measure been used effectively in practice? Does it have potential for working well with other indicators?*** |

In addition to the above framework, we followed the Donabedian paradigm of structure, process, and outcome[21, 22] to categorize current (HCUP I) and candidate (HCUP II) QIs. For example, potentially inappropriate utilization falls into the category of process, while in-hospital mortality, adverse events, and complication rates represent outcome measures.

We considered three broad audiences for the quality measures: health care providers and managers, who would use the quality measures to assist in initiatives to improve quality; public health policy-makers, who would use the information from indicators to target public health interventions; and health care purchasers and consumers, who would potentially use the measures to guide decisions about health policies and providers. Because of the limitations of quality indicators derived based on administrative data, we focus primarily on applications oriented to "screening for potential quality

problems." For the purpose of the Evaluation Framework, indicators must at least pass tests indicating that they are appropriate for the use of screening. In the rest of this section, we provide a more detailed explanation of each part of the Evaluation Framework, considering these three audiences wherever differences have been noted in the literature.

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

This question considers the degree to which potential users view the quality indicator as important and informative. There are two parts to this question: Does the indicator relate to an aspect of health care that users regard as important? And does performance on the measure credibly indicate high quality care? Obviously, face validity will be influenced by how well the indicator performs in the other areas covered in our Evaluation Framework. Clinicians tend to distrust outcome measures because of concerns over the adequacy of risk adjustment and the multiple factors beyond providers' control that contribute to poor outcomes. Other critics add that outcome measures suffer from imprecision (with random noise outweighing provider differences) and important selection biases (e.g., due to variations in admitting practices). Addressing this issue at the outset serves as a point of reference for the findings of our literature review and empirical analysis.

Broadly speaking, consumers, health care payers, regulators, and public health officials are likely to be most interested in measures based on outcomes that are relatively frequent, costly, and/or have serious implications for an individual's health. In addition, there should be reason to believe that the outcome may be (at least somewhat) under providers' control (in other words, controlled trials or well-designed cohort studies have shown that specific diagnostic or therapeutic modalities may reduce its frequency or severity). Outcome measures might include operative mortality rates, or mortality after hospitalization with serious acute illnesses such as a heart attack. These measures seem most intuitive, since they assess the main outcomes that medical treatments are intended to affect. Perhaps surprisingly, however, reports of hospital mortality rates appear to have little effect on where patients seek their care.[23, 24] One reason may be that many patients describe difficulty in interpreting indicators involving mortality and morbidity rates, and consequently view them as unhelpful.[25] Another reason may be that providers prefer measures of process, particularly if there is reason to believe (generally from randomized controlled trials) that certain processes truly lead to better patient outcomes. Patients appear to prefer reports of other patients' satisfaction with care, and especially informal recommendations from family, friends, and their own physicians.[25] Thus, developing indicators with high face validity for patients may require active participation from patients,[26, 27] targeting aspects of care identified as important in patient surveys,[28-35] or taking additional steps to enhance provider perceptions about the validity of outcome measures.

Many providers view outcome-based QIs with considerable skepticism.[36] For most outcomes, the impacts of random variation[37-42] and patient factors beyond providers' control often overwhelm differences attributable to provider quality.

Consequently, providers tend to support measures of quality based on processes of care that have been documented in clinical trials to lead to better health outcomes in relatively broad groups of patients[43-46] – for example, the processes of acute MI care measured in the Cooperative Cardiovascular Project.[47, 48] Such process measures focus precisely on the aspects of care under providers' control. As long as the process measures are based on evidence of effectiveness, they serve as useful proxies for outcome measures that would otherwise be difficult to observe or measure. For example, when using inpatient discharge data only, it is not possible to ascertain out-of-hospital mortality. In general, process measures are not as noisy as outcome measures because they are less subject to random variation. They also suggest specific steps that providers may take to improve outcomes or reduce costs – even if such outcome improvements are difficult to document at the level of particular providers.

The relationship between some structural quality measures and important outcomes has been well-documented, although some concerns remain about the interpretation of the measures.[21, 22, 49, 50] These measures include measures of hospital volume for volume-sensitive conditions,[51-66] technological capabilities (e.g., ability to perform certain intensive procedures like coronary angioplasty),[67-73] and teaching status.[73-79] All of these measures have limited face validity, because they are widely acknowledged to be weak surrogates for true quality of care.[80] For example, many low-volume hospitals have been shown to achieve excellent outcomes, whereas many high-volume hospitals have surprisingly poor outcomes.

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

The impact of chance on apparent provider or community health system performance must be considered. Unobserved patient and environmental factors may result in substantial differences in performance among providers in the absence of true quality differences. Moreover, the same providers may appear to change from year to year, in the absence of changes in the care they deliver. Thus, using "raw" quality data will often result in poorly reproducible, or imprecise, measurements, giving an incorrect impression of provider quality.

An extensive literature on the importance of random variations in quality measures now exists.[37, 39-42, 81-86] In general, random variation is most problematic when there are relatively few observations per provider, when adverse outcome rates are relatively low, and when providers have little control over patient outcomes or when variation in important processes of care is minimal. If a large number of patient factors that are difficult to observe influence whether or not a patient has an adverse outcome, it may be difficult to separate the "quality signal" from the surrounding noise. We review the evidence on the precision of each of the evaluated QIs. Empirical methods can be used to assess both the importance of sample size and the importance of provider effects (versus patient and area effects) in explaining observed variation in the measure. But this is not entirely a statistical question, and considerations of mechanisms and concerns related to face validity can also be helpful in assessing the precision of a measure. For example, if better hospitals invariably admit sicker patients, then the apparent variation in a measure at the hospital level will be significantly less than the true variation (see the

discussion of unbiasedness below). In such a case, other sources of evidence suggesting that a measure is valid or that such bias exists can be helpful in assessing the quality measure. Our literature review encompasses both empirical and other sources of evidence on measure precision, and our empirical analysis presents systematic evidence on the extent of provider-level or area-level variation in each quality measure.

Statistical techniques can account for random variations in provider performance by estimating the extent to which variation across providers appears to be clustered at the provider level, versus the extent to which it can be explained by patient and area effects.[86-89] Under reasonable statistical assumptions, the resulting estimates of the extent to which quality truly varies at the provider or area level can be used to "smooth" or "shrink" estimates of the quality of specific providers or areas. The methods are Bayesian: the data used to construct the quality measures are used to update a "prior" distribution of provider quality estimates, so that the "posterior" or smoothed estimate of a provider's (or area's) quality is a best guess, reflecting the apparent patient- and provider-level (or area-level) variance of measure performance.

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

A QI may exhibit precision, but nonetheless yield inaccurate results due to systematic measurement biases. Extensive research has documented the importance of *selection problems* in interpreting many quality measures, especially measures related to mortality.[90-94] Such biases may have two basic forms: differences in admitting practices between two hospitals produce non-random samples from the same underlying patient population (selection biases) or the patient populations may in fact contain different case-mixes. Selection effects presumably exert a greater influence on measures involving elective admissions and procedures, for which physician admission and treatment practice styles show marked variation.[74, 75] Nonetheless, selection problems exist even for conditions involving urgent "non-discretionary" admissions, likely due to modest practice variation [95] and non-random distribution of patient characteristics across hospital catchment areas.[77] The attention of researchers and quality analysts has focused on developing valid models to adjust for patient factors, especially when comparing hospital mortality.[90, 92]

The principal statistical approach to address concerns about bias is risk adjustment.[12, 15, 78, 79, 96-102] Numerous risk adjustment instruments currently exist,[15] but current methods are far from perfect.[15, 16] In general, risk adjustment methods are based on data drawn from two sources, administrative data and medical chart reviews.[12] Previous studies suggest that administrative data have at least two major limitations. First, coding errors and variations are common: some diagnoses are frequently entered with errors and with some inconsistency across hospitals.[103-105] Factors affecting the accuracy of these codes include restrictions on the number of secondary diagnoses permitted,[106, 107] as well as systematic biases in documentation and coding practices introduced by awareness that risk-adjustment and reimbursement are related to the presence of particular complications.[108-111]

Second, most administrative data sources do not distinguish disorders that can be in-hospital complications from pre-existing comorbidities.[12, 112] To the extent that diagnoses such as shock and pulmonary edema may result from poor quality of care, their incorporation in prediction models may bias estimates of expected mortality, and even favor hospitals whose care results in more complications. One proprietary risk-adjustment system has been shown to be significantly biased by its inclusion of conditions that actually developed after admission,[113] but this study was limited to one condition (acute MI) and its conclusions are somewhat controversial.[114] In another study, estimates of mortality differences between municipal and voluntary hospitals in New York City were substantially affected by whether potential complications were excluded from risk-adjustment.[79] New York and California have recently added a "6$^{th}$ digit" to ICD-9 codes to distinguish secondary diagnoses present at admission from those that developed during hospitalization. This refinement may allow valid comparisons of risk-adjusted mortality using administrative data for certain conditions,[115] although the accuracy of the "6$^{th}$ digit" has not been established.

Clinically-based risk adjustment systems supplement hospital discharge data with information available from medical records. Because exact clinical criteria can be specified for determining whether a diagnosis is present, coding errors are diminished. In addition, complications can be distinguished from comorbidities focusing on whether the diagnosis was present at admission.[15] Because the number of clinical variables that may potentially influence outcomes is small, and because these factors differ to some extent across diseases and procedures, progress in risk-adjustment has generally occurred by focusing on patients with specific conditions. Thus, sophisticated chart-based risk adjustment methods have been developed and applied for interpreting mortality rates for patients undergoing cardiac surgery[54, 116-118] and interventional cardiology procedures,[101] critically ill patients,[119, 120] patients undergoing general surgery, and medical patients with acute myocardial infarction,[47] community-acquired pneumonia, [121] and upper gastrointestinal hemorrhage.[122]

However, chart-based risk adjustment methods are not without their own limitations. First, especially for severely ill patients and those who die soon after admission – some of the most important patients for computing many quality measures – complete diagnosis information may not have been ascertained prior to death, and therefore would not be in the patient's medical record. Important observations might be missing for such patients, resulting in biased estimates in the risk-adjusted model. Second, medical chart reviews are very costly, and so routine collection of detailed risk information is not always feasible. As a result, the impact of chart-based risk adjustment may vary across measures. For some measures, its impact is modest and does not substantially alter relative rankings of providers.[19, 123-125] For others, it is much more important.[9-13, 18, 112] Of course, because all risk adjustment methods generally leave a substantial amount of outcome variation unexplained, it is possible that unmeasured differences in patient mix are important even in the most detailed chart-based measures.

For each quality measure, our report reviews the evidence on whether important systematic differences in patient mix exist at the provider and community level, and whether various risk adjustments significantly alter the quality measure for particular providers. We distinguish between risk adjustment methods that rely only on administrative data and have been validated with clinical data, and those that are not

validated. Risk adjustment methods requiring clinical data cannot be applied to the HCUP data, and therefore are not covered in this report. Our empirical analysis then assesses whether a common approach to risk adjustment using administrative data – the All Patient Refined Diagnosis Related Groups (APR-DRG) system developed by 3M™ – significantly alters the quality measure for specific providers. We emphasize impact on *relative* measures of performance – whether risk adjustment affects which hospitals are regarded as high- or low-quality – rather than *absolute* measures of performance – whether risk adjustment affects a hospital's quantitative performance on the quality measure. As noted above, we recognize that this system is not ideal, because it provides only four severity levels within each base APR-DRG, omits important physiologic and functional predictors, and potentially misadjusts for iatrogenic complications.

A remaining methodological issue concerns the appropriateness of adjusting for certain "risk factors."[126-135] For example, "Do Not Resuscitate" status may be associated with differences in care that not only reflect patient preferences (e.g., less use of intensive treatments) but also true differences in quality of care (e.g., inadequate physician visits), resulting in increased complications that would result in a "Do Not Resuscitate" order, and increased mortality.[136] Importantly, the prevalence of patients with DNR status may vary nonrandomly between hospitals, with large referral centers having greater percentages of patients seeking (and receiving) aggressive medical care.[137]

Adjusting for race implies that patients of different races respond differently to the same treatments, when patients of different races may actually receive different treatments. A substantial literature documents systematic differences in the care delivered to patients by race and gender.[125, 138-144] For example, African-American diabetics undergo limb amputations more often than do diabetics of other races.[145] Thus, we note wherever possible if review of the literature indicates particularly large differences in a quality measure by race or gender. Some gender or race differences may be due to either patient preference, or physiological differences that would be appropriate to include in a risk adjustment model. In other cases differences denote lower quality care, and in this case race and gender should not be included in the risk adjustment model. Where applicable this is noted in the literature review.

### 4. Construct validity: does the indicator perform well in identifying providers with quality problems?

Ideally, a hospital will perform well on a quality measure if and only if it does not have a significant quality problem, and will perform poorly if and only if it does. In practice, of course, no measure performs that well. Our analyses of noise and bias problems with each measure are intended to assess two of the principal reasons why a hospital might appear relatively good or bad (or not appear so) when it really is not (or really is). Detecting quality problems is further complicated by the fact that adverse outcomes are often the result of the course of an illness, rather than an indication of a quality problem at a hospital. Formally, one would like to know the sensitivity and

specificity of a quality measure, or at least the positive predictive value (PPV) of a quality measure for detecting a true hospital quality problem.[2]

When available, for each measure, we reviewed any existing literature on its sensitivity or PPV for true provider quality problems. In most cases, however, we found no true gold standard, or ideal measure of quality. We therefore tested construct validity – i.e., our construct is that different measures of quality, on the same patients, should be related to each other at the provider level, even if it isn't always clear which measure is better. It may be easier to ask "is the indicator correlated with other, accepted measures of quality at the provider level?" rather than "does the indicator perform well in identifying providers with quality problems?" For example, studies have validated survey rankings of "best" hospitals by examining the relation with actual process and outcome measures for AMI,[146] and peer review failure rates with HCFA risk-adjusted mortality rates.[147]

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

Ideally, when quality measures are used to guide quality improvement initiatives or reward good providers, the best way for a provider to perform well on the measure is to provide high-quality care. Unfortunately, many quality indicators appear to at least leave open the possibility of improving *measured* performance without improving *true* quality of care.

In measures that are risk-adjusted, measured performance can be improved by "upcoding" – including more comorbid diagnoses in order to increase apparent severity of illness.[86, 111] Systematic biases in diagnostic codes were observed after the introduction of the Prospective Payment System,[108-110] and may also explain much of the apparent reduction in adjusted mortality attributed to the Cardiac Surgery Reporting System in New York.[111] The extent to which upcoding is a problem probably increases with the ambiguity of the specific data element, and decreases when auditing programs maximize the reliability and validity of submitted data. In recent years, an aggressive auditing program has significantly reduced the extent to which comorbidities not substantiated by the medical chart are recorded for Medicare patients, leading some analysts to conclude that "upcoding" is no longer as substantial of a problem for Medicare patients.[148] However, such audit standards have generally not been imposed on the state discharge databases used in the HCUP project. In our review, we note indicators for which risk adjustment appears to be important, and thus upcoding is a potentially important problem.

Indicators capturing patient morbidity, such as adverse events and complications, must overcome a reporting bias in the reverse direction - i.e., towards under-reporting. With some exceptions, most hospitals in most states rely on voluntary incident reporting for adverse events. Such methods are known to detect only a fraction of true adverse

[2] The PPV represents that the chance that a positive test result reflects a "true positive." It combines the properties of the test itself (e.g., sensitivity and specificity for detecting quality problems) with the prevalence of true quality problems in the target population.

drug events (ADEs).[149] The Institute of Medicine has recently recommended mandatory reporting systems for adverse events emanating from certain egregious errors. [150] However, the JCAHO's sentinel reporting system tracks many of these same errors (e.g., operating on the wrong patient or body part, suicide or rape of an inpatient), and it was received very negatively by hospitals, despite being a voluntary system. Thus, the degree to which mandatory reporting requirements alleviate or exacerbate reporting bias for adverse events remains to be seen. In addition, high-quality hospitals with sophisticated error detection systems may report errors more frequently, leading to high apparent complication rates in hospitals that may have superior quality in other dimensions.[151-153]

Perverse incentives may arise from the criteria used to define or identify the target patient population. For instance, restricting mortality measures to inpatient deaths potentially allows hospitals to lower their mortality rates simply by discharging patients to die at home or in other institutions.[106, 115, 154, 155] Measures of surgical site infections[156, 157] and other complications of hospital care[158] that only capture in-hospital events will similarly reward hospitals that merely reduce length of stay by discharging or transferring high-risk cases. Early concerns that surgeons in New York avoided operating on high-risk patients[159] may have proved unfounded,[160, 161] though this issue remains unsettled.[162] In general, the incentive for providers to avoid treating sicker patients remains a significant concern for outcome-based quality measures.[86]

We reviewed the available evidence on each of these possible undesirable responses to the use of each quality measure. For the most part evidence was lacking on responses to indicators, particularly since many of the proposed indicators have not been subjected to public reporting. We noted potential responses when appropriate.

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

While important problems exist with many specific applications of HCUP and other quality indicators, they have been applied in a range of settings. As noted in the section on face validity, these applications broadly include initiatives to improve provider quality and initiatives to provide quality-related information to providers and consumers. For each quality indicator, we reviewed studies describing its use in these activities. However, we did not conduct a thorough review of the non-peer reviewed literature. Therefore, indicators may have been adopted, and may continue to be used, by many provider organizations or government agencies without our knowledge.

A recent systematic review more comprehensively summarizes the literature on the impact of performance reports on consumers, providers, and purchasers.[163] Useful and accurate information on quality remains a desirable goal for consumers and providers alike. The interest in quality and the resulting data and research has had some impact on the field of health services research. For instance, the HCUP project has provided a valuable resource for a number of studies in health services research.[3, 133-135, 164-177]

# 2.C. Literature Review of Quality Indicators

We conducted a literature review to identify quality indicators reported as such, and potential quality measures. The result of this first stage was a comprehensive list of measures that could be defined based on routinely collected hospital discharge data. In the second phase, we searched the literature for further evidence on these indicators to provide information on their suitability for the HCUP II QI set. This second phase resulted in a comprehensive bibliography for each indicator. In addition, a sub-set of the entire indicator list was selected for detailed review using specific evaluation criteria (described in Section 2.B). The entire process for this systematic review of the literature is diagrammed in Flow Diagram 1, and described in the following sections.

## *Phase 1: Identification of Indicators*

**Step 1: Selecting the articles.** To locate literature pertaining to quality indicators, we conducted a strategic literature search using the Medline database. We compared over 30 search strategies using Medical Subject Headings (MeSH) based on their ability to retrieve a set of key articles known to the project team. Successful combinations of MeSH term searches returned all the key articles. The final MeSH terms used were "hospital, statistic and methods" and "quality indicators." Articles were also limited to those published in 1994 or later. Articles prior to 1994 had been reviewed for HCUP I development. This search returned the highest number of known key articles in the most concise manner, returning approximately 2,600 articles.

Articles were screened using the titles and abstracts for preliminary abstraction. To qualify for preliminary abstraction the articles must have described a potential indicator or quality relationship that could be adequately defined using administrative data, and be generalizable to a national data set. For the purpose of this study, a quality indicator was defined as an explicit measure (defined by the developer) of some aspect of healthcare quality. Some literature defines only a quality relationship, in that the articles expounds on a process or structural aspect of a healthcare provider that is related to better outcomes. However, the author does not specifically define or recommend that the relationship be used as a quality measure. In this case, the article only describes a quality relationship, not a quality indicator. Only 181 articles met the criteria for preliminary abstraction (See Appendix 3 for a list of abstracted articles). This reflects the small number of quality indicators with published formal peer-reviewed evaluations.

**Step 2: Preliminary abstraction.** The preliminary round was designed to screen articles for applicability and quality, to obtain and assess the clinical rationale of the indicators, and to identify those articles with enough detail for a more comprehensive abstraction. Nine abstractors participated in this phase. Five of these abstractors were medical doctors with health services research training. The remaining four abstractors were familiar with the project and the literature, and included a project manager (JG), the research coordinator (SD) and two undergraduate research assistants (PC, MW).

The articles were sorted into clinical groupings. The research coordinator rated these clinical groupings according to the amount of clinical knowledge required to

abstract the articles. Those requiring the most clinical knowledge were assigned to physicians, while those requiring the least clinical knowledge were assigned to the undergraduate research assistants. Abstractors selected clinical groupings that were of interest or that corresponded to their clinical specialties.

Abstractors recorded information about each article on a one-page abstraction form(see Appendix 4, Data Collection Forms, Preliminary Abstraction Form). Information coded included:

1. Indicator type (i.e. mortality, readmission, potentially overused procedures)
2. Clinical domain (i.e. medical, surgical, obstetric, pediatric, and psychiatric)
3. Measure category (i.e. structure, process, proxy-outcome, and outcome)
4. Clinical rationale for the indicators.
5. Use of longitudinal data.
6. Use of data beyond hospital discharge data.
7. Strengths and weaknesses identified by the author.
8. Strengths and weaknesses not identified by the author.

Each abstraction form was reviewed by the research coordinator for quality of the abstraction, and for accuracy of the coding. All data were then entered into a Microsoft Access database.

**Step 3: Full abstraction.** The purpose of the full abstraction phase was to identify potential indicators for the HCUP II QI set, and to assess the evidence for validity of existing indicators. To accomplish this, only articles that described an indicator in conjunction with specific and comprehensive information on its validity were fully abstracted. Four of the original abstractors participated in this phase of the abstraction. Three of these abstractors were medical doctors, the fourth a master's level research coordinator.

Each of the articles for preliminary abstraction, and the corresponding abstraction form was reviewed by both the research coordinator and the project manager independently. To qualify for full abstraction the articles needed to meet the previously noted criteria, and the following criteria:

1. Define a quality indicator, as opposed to only a relationship that was not formulated or explicitly proposed as a measurement tool.
2. Discuss a novel indicator, as opposed to indicators defined elsewhere and used in the article only to discuss its relationship with another variable (i.e. SES, race, urbanization).
3. Define an indicator based on administrative data only.

Only 27 articles met these formal criteria (see Appendix 3). This highlights an important aspect of the literature on quality indicators: most indicators are based on published clinical literature to identify important patient and provider characteristics and processes of care for specific clinical conditions; there is also a substantial literature on technical aspects such as severity adjustment, coding and data collection. It should be noted that, while only 27 articles qualified for formal abstraction, these are not the only useful articles. Many articles provide important information about quality measurement.

However, few quality indicators are specifically defined, evaluated, and reported in the literature besides descriptive information on the process of development (the Complication Screening Program is a noteworthy and laudable exception, in that it has been extensively validated in the published literature, mostly by the developers). This evidence report will be an important contribution to the paucity of literature on indicator validation.

An abstraction form was filled out for each indicator defined in an article. The abstraction form coded the following information(see Appendix 4, Data Collection Instruments, Indicator Abstraction Form):

1. All the information coded in the preliminary abstraction form.
2. Measure administrative information (i.e. developer, measure set name, year published).
3. Level of care (primary (prevention), secondary (screening or early detection) or tertiary (treatment to prevent mortality/morbidity)).
4. Scoring method (i.e. rate, ratio, mean, proportion).
5. A priori suggested quality standard (i.e. accepted benchmark, external comparison, and internal comparison).
6. Indicator definition (Numerator, denominator statements, inclusions, and exclusions).
7. Extent of prior use.
8. Current status (i.e. measure defined, pilot tested, implemented, discontinued).
9. Scientific support for measure (i.e. published guidelines, clinician panel, literature review, revision of pre-existing instruments, theory only).
10. Other essential references for the measure.
11. Validity testing.
12. Risk adjustment.

If the measure included risk adjustment, a separate form for the risk adjustment method was filled out (see Appendix 4, Data Collection Instruments, Risk Adjustment Abstraction Form). This included:

1. Method administrative information.
2. Adjustment rationale.
3. Classification or analytic approach (i.e. stratification, logistic or linear regression)
4. System development method (i.e. logistic regression, score based on empirical model, a priori/clinical judgement).
5. Published performance for discrimination and calibration.
6. Use of co-morbidities, severity of illness, or patients demographics.
7. Use of longitudinal data, or additional data sources beyond discharge data.
8. Extent of current use.
9. Other essential references for the method.
10. Abstractor comments.

The abstraction forms were reviewed by the research coordinator and entered into a Microsoft Access database.

**Parallel Step: Supplementing literature review using other sources**. Because the literature in this area is not the primary source for reporting the use of quality indicators, we compiled a list of suitable indicators from a variety of sources. As previously noted, the phone interviews with project advisors led to information on some indicators. In addition, we searched the world wide web sites of known organizations using quality indicators, the CONQUEST database, National Library of Healthcare Indicators (NLHI), developed by the Joint Commission on Accreditation of Healthcare Organizations (JCAHO), and a list of ORYX approved indicators provided by the JCAHO. Indicators that could be defined using administrative data were recorded in an indicator database.

**Breakdown of indicators by primary source.** During Phase 1, we found that no one literature search was sufficiently sensitive for the purpose of identifying either quality indicators or quality relationships. In addition, we found relatively little literature defining quality indicators. In order to be confident that we identified a large percentage of the quality indicators in use we also performed the parallel step above. This parallel step led us to websites, organizations, and additional literature describing quality indicators. In general, most volume, utilization, and ACSC indicators have been described primarily in the literature. On the other hand, the primary sources for most mortality and length of stay indicators were current users or databases of indicators. However, many indicators found in the literature, were also reported by organizations, and vice versa. Thus, it is difficult to delineate which indicators were derived only from the literature, and which were derived from the parallel step described above.

## *Phase 2: Evaluation of Indicators*

The result of Phase 1 was a list of potential indicators with varied information on each depending on the source. Since each indicator relates to an area that potentially screens for quality issues, we developed a structured evaluation framework to determine measurement performance. A series of literature searches were then conducted to assemble the available scientific evidence on the quality relationship each indicator purported to measure. Due to limited resources, we could not review all of the indicators we identified in Phase 1, and therefore selected some for detailed review using the evaluation framework. The criteria used to select these indicators are described later.

**Step 1. Evaluation framework development**. As described previously, a structured evaluation of each indicator was developed and applied to assess indicator performance in six areas:

1) Face validity
2) Precision
3) Minimum Bias
4) Construct Validity
5) Fosters Real Quality Improvement
6) Prior use

**Step 2. Identification of the evidence.** We searched the literature for evidence in each of the six areas of indicator performance described above, and in the clinical areas addressed by the indicators. The search strategy used for Phase 2 began with extensive electronic searching of MEDLINE,[178] PsycINFO,[179] and the Cochrane Library.[180] [3] Thus, in contrast to conducting systematic reviews of purely clinical topics, we reasoned that the European literature not captured in the Medline database or Cochrane Library would almost certainly represent studies of questionable relevance to the US health system.

The extensive electronic search strategy involved combinations of MeSH terms and keywords pertaining to clinical conditions, study methodology, and quality measurement (see Figure 2).

**Figure 2. Example search strings**

Clinical:

I. (Laminectomy OR Intervertebral Disk
II. Arthroplasty, Replacement,
III. (Hip prosthesis) OR (Arthroplasty, Replacement, Hip) OR (Hip
IV. (Intensive Care) OR (Intensive Care Units) OR (Critical Care) OR illness)
V. Myocardial
VI. Surgery Department,
VII. Postoperative
VIII. Cross Infection ,Pneumonia/prevention & control
Artificial/adverse effects, Respiratory Tract Infections/prevention &
IX. (Cross Infection/etiology) OR (Infection Control) OR (Antibiotic
X. Clinical
XI. (Gastrointestinal Hemorrhage/prevention & control) OR
Hemorrhage/mortalit
XII. Respiration, Artificial/methods •Respiratory
XIII. Bacteremia/prevention &
XIV. Catheterization, Central Venous,Catheters,
effects/microbiology, Equipment

Methodological:

1. (Quality Assurance, Health Care) OR (Outcome assessment, health care) OR (Process assessment, healthcare) OR (Quality indicators in health care)
2. (Severity of Illness Index) OR (Logistic Models) OR (Proportional Hazards Models) OR (Severity of Illness Index) OR (Models, Statistical)
3. (Hospital Mortality) OR (Survival analysis) OR (Treatment Outcome)
4. Physician's Practice Patterns
5. Hospitals/standards
1. 6. Hospitals, Group Practice/statistics & numerical data/standards
7. Patient Admission
8. Utilization Review/statistics & numerical data
9. Practice guidelines, guideline adherence
10. Predictive Value of Tests
I. Selection Bias
II. Residence Characteristics
III. (Evaluation Studies) OR (Cohort Studies) OR (Comparative Study) OR (Retrospective studies) OR (Prospective studies) OR (Follow-Up Studies) OR (Randomized controlled trials)

[3] We chose not to search EMBASE[181] on the grounds that the studies of quality measurement necessarily must take into account the particular health care system involved.

**Figure 3. Example search**

**Example Search - Mortality Following Stroke**

| Medline Search String | Number of References Retrieved |
|---|---|
| 1) Cerebrovascular Disorders [MeSH Terms] | 47,264 |
| 2) Epidemiologic Studies [MeSH Terms] OR Clinical Trials [MeSH Terms] | 32,630 |
| 3) Search mortality [MeSH Terms] OR prognosis [MeSH Terms] | 18,460 |
| 4) #1 AND #2 AND #3 | 2,410 |
| 5) #4 AND stroke [ti] | 524 |
| 6) Quality of Health Care [MeSH Terms] | 852,714 |
| 7) #1 AND #2 AND (#3 OR #6) | 1,988 |
| 8) Reproducibility of Results [MeSH Terms] OR Sensitivity and Specificity [MeSH Terms] | 110,384 |
| 9) records [MeSH Terms] OR Hospitalization [MeSH Terms] | 55,739 |
| 10) #8 AND #9 | 3,835 |
| 11) #1 AND #10 | 106 |

All searches included Limits: Publication Date from 1990 to 2000, and Language English.

The results of searches 5 and 11 were scanned (titles and abstracts) to pull relevant studies, and the bibliographies of these studies were hand-searched for additional references.

Additional literature searches were conducted using specific measure sets as "keywords". These included "Maryland Quality Indicators Project," "HEDIS and low birth weight, or cesarean section, or frequency, or inpatient utilization," "IMSystem," "DEMPAQ," and "Complications Screening Program."

We also searched the bibliographies of key articles, and hand searched the Tables of Contents of general medical journals, as well as journals focusing in health services research or in quality measurement. This list of journals included *Medical Care, Health*

*Services Research, Health Affairs, Milbank Quarterly, Inquiry, International Journal for Quality in Healthcare,* and *the Joint Commission Journal on Quality Improvement.*

These literature searches and on-line screening for relevancy retrieved over 2000 additional articles, which were added to the project database. These articles were used for evaluations of individual indicators. Those articles cited are listed in the references.

The use of medical literature databases likely eliminated much of the "gray literature" that may be applicable to this study. Given the limitations and scope of this study, we did not complete a formal search of the "gray literature" beyond that which was previously known by the project team, or resulted from our telephone interviews.

**Step 3. Selection of a sub-set of indicators.** Since there were too many indicators identified in Phase 1 (literature search and parallel steps) for detailed evaluation using the Evaluation Framework , we developed criteria to select a group for further evaluation. These criteria were intended to be top level evaluations of the face validity and precision of the indicators. We first selected a subset of indicators for preliminary empirical evaluation. To do this, first, the indicators related to complications were disqualified for this particular report since they will be included in an expansion to the report that will include patient safety indicators. Second, all of the current HCUP I indicators (except those related to complications of care) were selected for empirical evaluation. Third, we lowered the priority of clinical areas well covered by the current HCUP I indicator set, for example obstetrical indicators. Finally, a set of criteria for selection was applied to the remaining indicators.

The following were specific criteria for evaluation for all indicators:

- Indicator must be definable with HCUP data (i.e. uses only administrative data available in HCUP data set)
- Conditions that affect at least 1% of hospitalized patients and/or 20% of providers, as tested using the Nationwide Inpatient Sample Dataset
- Conditions that are the subject of public reporting, previous use or large dollar volume
- Clear relationship to quality apparent as evaluated by clinical judgement of health services researchers and medical doctors

In addition several specific criteria were noted for the indicator types:

- Volume:

1. Widely documented volume-outcome relationship
2. Recent evidence regarding volume-outcome relationship

- Utilization rates:

1. Condition must have an alternative surgical or medical therapy with lower/higher morbidity and/or mortality

- Ambulatory care sensitive conditions:

1. Differences in patient management practices for that condition
2. Existence of treatment guidelines, and evidence of failure to comply

- In-hospital mortality

1. Relatively homogenous group

When selecting between competing alternatives that met all the above criteria, we chose to evaluate clinical areas in depth rather than evaluating a large breadth of indicators. To do this we evaluated multiple aspects in one clinical domain (i.e., evaluations of CABG, PTCA, and AMI; stroke and carotid endarterectomy). In these clinical areas we evaluated at least two different types of indicators (i.e. mortality and utilization). The reasons for selection or elimination for each indicator are listed in Appendix 5.

The selected indicators were then evaluated empirically, using preliminary tests of precision. Those demonstrating adequate precision were then evaluated by a literature review (Phase 2), as well as further empirical analysis.

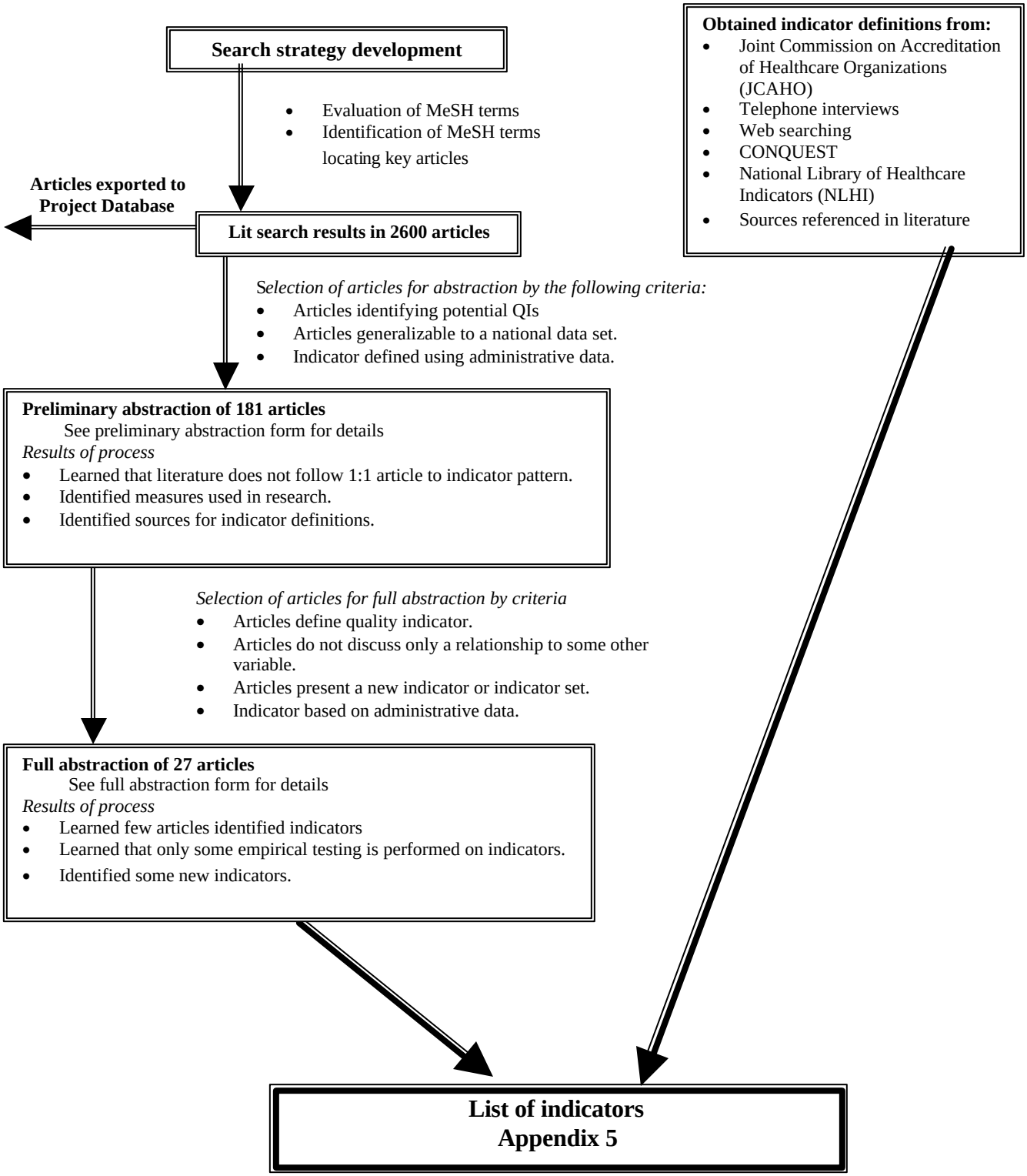
**Step 4. Evidence evaluation**. The abstracts from relevant articles for each indicator were reviewed and selected according to the following criteria:

1) The article addressed some aspect of the six areas of indicator performance.
2) The article was relevant to a national sample, rather than a local population.

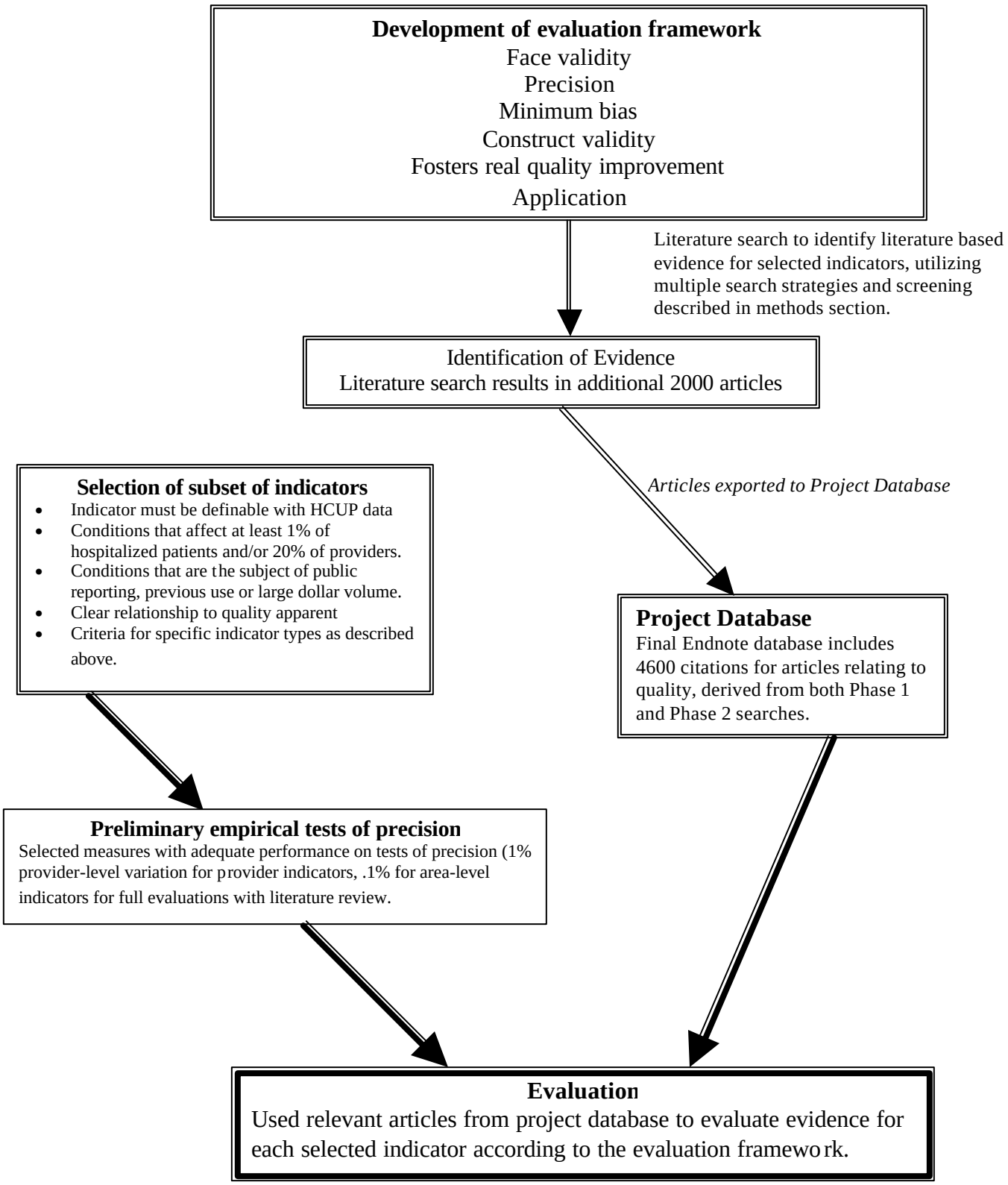
Based on this literature, a team member or clinician developed a draft write-up of the indicator following the evaluation framework (See Section 2.B. for a description of the evaluation framework).

# Flow Diagram 1. Literature review strategy

## *Phase 1. Identification of indicators*



***Phase 2. Evaluation of indicators***



# 2.D. Risk Adjustment of HCUP Quality Indicators

***Overview.*** "Raw" unadjusted measures of hospital or area performance for each indicator are simple means constructed from the HCUP discharge data and census population counts. Obviously, simple means do not account for differences in the indicators that are attributable to differences in patient mix across hospitals that are measured in the discharge data, or demographic differences across areas. In general, risk adjustment involves conducting a multivariate regression to adjust expected performance for these measured patient and population characteristics. Although complex, multivariate regression methods are the standard technique for risk-adjustment because they permit the simultaneous consideration of multiple patient characteristics, and interaction among those characteristics. The interpretation of the risk-adjusted estimate is straightforward: it is the value of the indicator expected at that hospital if the hospital had an 'average' patient case-mix.

This section contains the methods for our evaluation of risk adjustment systems, leading to our decision to use APR-DRGs. The purpose of this evaluation is to briefly outline the evidence regarding the use of risk adjustment systems for both evaluating potential bias in indicators, and for risk adjusting established indicators to compare provider performance. The first section discusses criteria used to evaluate the risk adjustment systems. Such criteria arise from both the literature-based evidence on risk adjustment systems, as well as user criteria obtained through the semi-structured telephone interviews. Second, we outline the methods used to implement APR-DRGs empirically in the HCUP II QI set. We describe our methods for risk-adjustment of the hospital level quality indicators. An analogous method was used for the area level quality indicators. However, in our area level indicators we account only for demographic differences.

## Risk Adjustment Literature Review Methods

Our literature review for risk adjustment of the HCUP Quality Indicators combined evaluation criteria common to evidence studies on the performance of risk adjustment systems with additional considerations of importance to the potential HCUP users. These considerations were determined through semi-structured interviews with users, discussed earlier in this report. In general, users viewed risk adjustment as an important component of the HCUP QIs refinement. State data organizations and agencies involved in reporting of hospital performance measures especially tended to view risk-adjustment as essential for the validity of the results and acceptance by participating hospitals. Concerns that patient severity differed systematically among providers, and that this difference might drive the performance results, was frequently mentioned as a reason for limited reporting and public release of the HCUP I QIs to date, especially for outcome-oriented measures like mortality following common elective procedures.

*Literature-Based Criteria for Evaluating Risk Adjustment Systems.* HCUP I users were concerned about the validity or performance of possible risk adjustment systems.

We assessed evidence on the performance of risk-adjustment systems from published reports using the following commonly applied criteria.[15, 16, 182]

1. *Classification and analytic approach.* Risk adjustment systems have been developed to predict complications, resource use, and mortality. Alternative analytic approaches included stratification (assigning individuals to mutually exclusive cells), logistic regression, or linear regression (calculating an expected level of medical utilization based on a statistical model). Methods based on logistic or linear statistical models are generally able to consider more dimensions of patient characteristics than stratification. Even more effective approaches might involve combining multivariate adjustment and stratification through propensity score methods[183] and accounting for the relationship between aspects of disease severity that are measured and those that are not.[184] However, no currently available risk adjustment systems are based on these analytic methods.

2. *System development method.* Risk adjustment classifications may be based either on an empirical model clinical judgment, or some combination. For example, an assessment of whether two heart attack patients are expected to have similar outcomes can be based on statistical tests or clinical expertise or both.[15]

3. *Feasibility.* Feasibility is largely determined by the data requirements of the risk-adjustment method. We reviewed whether a system required hospital data elements other than those found on the discharge abstract (e.g., data from medical charts or laboratory data) or non-hospital data (e.g., outpatient hospital or physician data). We also evaluated whether the method was likely to be enhanced with discharge data that included a unique patient identifier, so that risk adjusters could be developed based on data from multiple hospitalizations or encounters. Because only a subset of the states participating in HCUP collect supplementary data beyond discharge abstracts and/or unique patient identifiers for use in longitudinal analyses, we did not select a risk adjustment system that depended on such information.

4. *Empirical performance: discrimination*. A critical aspect of the performance of a risk-adjustment model is the extent to which the model predicts a higher probability of an event for patients who actually experience the event. The statistical test of discrimination is generally expressed as a C-statistic or $R^2$ (how much of the variation in the patient level data the model explains). In general, systems that discriminate more have the potential to influence QI measures more substantially. Many severity-adjustment systems were designed primarily to predict in subsequent periods (e.g. resource consumption next year). However, for purposes of evaluating QI performance, the estimation of concurrent risk is more important (i.e., differences in the likelihood of experiencing an outcome in the current time period). Ideally, discrimination would be assessed using an $R^2$ or other statistic of predicted variation that is computed on a separate data source from the one used to develop the model, to avoid 'over-fitting' (i.e., the model might appear do well in part because it explains nonsystematic variations in the data used to develop it).

5. *Empirical performance: calibration.* Calibration is a measure of whether the mean of the predicted outcomes equals the mean of the actual outcomes for the entire population and for population subgroups. The statistical test is often expressed as a Chi-square or 'goodness-of-fit' for the equivalence of means of population subgroups. Even if the severity-adjustment system does not predict well at the level of individuals, it may predict well at the aggregate (group) level of, say, women, 70-74 years of age. Over-fitting will be an issue here as well, unless a different data source is used to validate the model than was used to estimate the model.

Not many risk-adjustment systems have been evaluated in published reports using all of these criteria, nor have they been evaluated using consistent data sources. These limitations of the literature on risk adjustment complicate comparisons of risk adjustment systems based on performance criteria. In the end, the user specified criteria determined a narrow set of potential risk adjustment systems to consider. The performance criteria delineated between these potential systems, and informed our own empirical evaluation of the impact of risk adjustment on the assessment of provider and area quality.

*User-Specified Criteria for Evaluating Risk Adjustment Systems.* Evidence on the performance of a risk adjustment system is a primary consideration for HCUP users, and is essential to the validity of reported performance measures. However, users also cited other factors as potentially important determinants of the acceptance of HCUP QIs reporting by hospitals, state regulators and state legislatures, and other potential consumers of hospital performance data. These factors included the following:

1. *"Open" systems preferable to "black box" systems.* Although there was no specific prohibition against using proprietary systems vs. systems in the public domain, there was a preference for using "open" systems where the risk adjustment logic was published and available for scrutiny by interested parties.

2. *Data collection costs minimized, and well-justified*. The widespread recognition that data collection was costly for hospitals meant that any risk-adjustment system that would be imposed on hospitals had to justify the cost of data collection, by documenting that the additional information led to substantially different and more accurate inferences about performance. At least one state had stopped using a risk adjustment system that required medical chart review because the high cost of implementation was not considered worth the efficiency gained from improved accuracy.

3. *Multiple-use coding system*. Some risk adjustment systems were designed to categorize patients according to expected resource use, defined either as charges or length of stay, while others were designed to categorize patients according to expected health outcomes, including mortality and complications. For example, several states calculated and reported mortality rates by diagnosis-related group (DRG). These users generally believed that a risk-adjustment system for health outcomes based on discharge records that relied on the same diagnostic groups used for reimbursement was more likely to be accurate than a system that relied

on codes used for quality and health outcome comparisons only, since there would be less financial and audit incentives to record codes accurately for the latter. Thus, coding systems that affected reimbursement for at least some patients were likely to capture diagnoses and procedures reported in medical charts. One potentially important limitation of relying on codes that are also used for payment is that changes in reimbursement-related coding practices (e.g., as a result of tighter Medicare rules implemented in 1996) may alter apparent severity. However, because of the financial implications of changes in coding practices, any significant changes are likely to be identified and reported by payers, and so can be considered in interpreting variations and trends in reported quality measures.

4. *Official recognition.* Many users indicated that systems that had been supported or otherwise recognized by government agencies such as AHRQ were preferable to other systems, because such support facilitated acceptance by legislative and hospital groups. Adoption of the HCUP I QIs themselves were often justified in part by their sponsorship by the AHRQ. State agencies, especially those from smaller states, often cited the lack of staff resources and expertise needed to make independent evaluations of competing indicator sets and risk adjustment methods.

## Risk Adjustment Empirical Methods

We used the APR-DRG system, with severity and risk of mortality classifications, in two ways:

- to evaluate the impact of measured differences in patient severity on the relative performance of hospitals and areas, by comparing QI measures with and without risk adjustment; and

- to risk-adjust the hospital- and area-specific measures.

As our previous review noted, the available literature on the impact of risk adjustment on indicator performance is limited, but suggests that at least in some cases different systems may give different results. Problems of incomplete or inconsistent coding across institutions are probably important contributing factors to the differences in results. Thus, definitive risk adjustment for some indicators may require detailed reviews of medical charts and additional data sources (charts may also be incomplete), just as definitive quality measures for many indicators may require additional sources of information. However, the importance of random variations in patients means that whatever risk adjustment and quality measurement system is chosen should be used in conjunction with statistical methods that seek to minimize other sources of noise and bias. Our empirical analysis is intended to illustrate the approach of combining risk adjustment with smoothing techniques, including suggestive evidence on the importance of risk adjustment for potential HCUP II QIs, using a risk adjustment system that can be implemented on discharge data by most HCUP users. We also supplement our empirical analysis with a review of the clinical literature, to identify additional clinical information

that is important to consider for certain indicators. In particular, our literature review highlights a few indicators where risk adjustment with additional clinical data has been shown to be particularly important, and where important differences in case mix seem less likely to be related to the secondary diagnoses used to risk-adjust discharge data.

This section describes how we implement risk-adjustment using patient demographics (age and sex) along with the APR-DRG classification system. The next section describes statistical methods we employ to account for additional sources of noise and bias not accounted for by observed patient characteristics. By applying these methods to all of the potential HCUP II QIs, we are able to evaluate the relative importance of both risk adjustment and smoothing in terms of the relative performance of hospitals (or areas) compared to the "raw" unadjusted QIs based on simple means from NIS discharge data. The simple means fail to account both for differences in the indicators that are attributable to systematic differences in measured and unmeasured patient mix across hospitals/areas that are measured in the discharge data, and for random variations in patient mix. We adopted a multivariate regression approach to adjust performance measures for measured differences in patient mix, which permits the inclusion of multiple patient demographic and severity characteristics.

Specifically, if we denote whether or not the event associated with a particular indicator $Y^k$ (k=1,…,K) was observed for a particular patient i at hospital/area j (j=1,…,J) in year t (t=1,…,T), then the regression to construct a risk-adjusted "raw" estimate a hospital/area's performance on each indicator can be written as:

(1) $$Y^k_{ijt} = M^k_{jt} + Z_{ijt}\ \Pi^k_t + \xi^k_{ijt}\ ,$$ where

$Y^k_{ijt}$ is the $k^{th}$ quality indicator for patient i discharged from hospital/area j in year t (i.e., whether or not the event associated with the indicator occurred on that discharge);
$M^k_{jt}$ is the "raw" adjusted measure for indicator k for hospital/area j in year t (i.e., the hospital/area "fixed effect" in the patient-level regression);
$Z_{ijt}$ is a vector of patient covariates for patient i discharged from hospital/area j in year t (i.e., the patient-level measures used as risk adjusters);
$\Pi^k_t$ is a vector of parameters in each year t, giving the effect of each patient risk adjuster on indicator k (i.e., the magnitude of the risk adjustment associated with each patient measure); and
$\varepsilon^k_{ijt}$ is the unexplained residual in this patient-level model.

The hospital/area specific intercept $M^k_{jt}$ is the "raw" adjusted measure of a hospital/area's performance on the indicator, holding patient covariates constant. In most of the empirical analysis that follows, the patient-level analysis is conducted using data from all hospital/areas. (The model shown implies that each hospital/area has data for all years, and within each year has data on all outcomes; however, this is not essential in order to apply risk adjustment methods.)

We estimated these patient-level regressions by linear ordinary least-squares (OLS). In general, the dependent variables in the regressions are dichotomous, which

raises the question of whether a method for binary dependent variables such as logit or probit estimation might be more appropriate. However, previous work by McClellan and Staiger has successfully used OLS regression for similar analyses of hospital/area differences in outcomes. In addition, estimating logit or probit models with hospital/area fixed effects cannot be done with standard methods; it requires computationally intensive conditional maximum likelihood methods that are not easily extended to multiple years and multiple measures.[185] A commonly-used "solution" to this problem is to estimate a logit model without hospital/area effects, and then to use the resulting predictions as estimates of the expected outcome. However, this method yields biased estimates and predictions of hospital performance. In contrast, it is easy to incorporate hospital/area fixed effects into OLS regression analysis, the resulting estimates are not biased, and the hospital/area fixed effects provide direct and easily-interpretable estimates of the outcome rate for a particular hospital/area measure in a particular year, holding constant all observed patient characteristics.

Of course, it is possible that a linear probability model is not the correct functional form. However, as in our earlier work, we specified a very flexible functional form, including full interactions among age and sex covariates as well as a full set of APR-DRG risk adjusters. In our sensitivity analyses for selected quality measures, this flexible linear probability model produced estimates of the effects of the risk adjusters that did not differ substantially from nonlinear (logit and probit) models. Another potential limitation of the OLS approach is that it may yield biased estimates of confidence intervals, because the errors of a linear probability model are necessarily heteroskedastic. Given the large sample sizes for the parameters estimated from these regressions (most indicators involve thousands of "denominator" discharges per year), such efficiency is not likely to be an important concern. Nevertheless, we estimated models using Weighted Least Squares[186] to account for heteroskedasticity, to see if our estimates were affected. We obtained very similar estimates of adjusted indicator performance.

Specifically, in addition to age, sex, and age*sex interactions as adjusters in our model, we also included the APR-DRG category for the admission and the APR-DRG constructed severity subclass (or risk-of-mortality subclass for mortality measures). APR-DRGs are a refinement of the DRGs used by the Health Care Financing Administration, with additional classifications for non-Medicare cases (e.g., neonates). The severity subclass evaluates the episode of care on a scale of 1 (minor) to 4 (extreme). In the APR-DRG Version 12, Severity of Illness is defined as the "extent of physiologic de-compensation or organ system loss of function." The APR-DRG severity of illness subclass was designed principally to predict resource use, particularly length-of-stay. As such, because this risk-adjustment system was not designed to predict utilization rates, for example, in our evaluation of each indicator we do not consider lack of impact of risk-adjustment to be evidence of lack of real bias. However, we do consider impact of risk-adjustment to be evidence of problems of potential bias. The literature review further informs us to potential sources of bias, and the prior use of each indicator may require collection of supplemental data for confounding clinical conditions. We discuss these issues further in the Results Chapter.

For each indicator, we excluded from our risk adjustment model the APR-DRG groupings in the Major Diagnostic Category (MDC) related to that indicator. The

groupings are either medical (based on diagnoses) or surgical (based on procedures), and we excluded groupings in the MDC of the same type. For example, for the Coronary Artery Bypass Graft rate indicator, we excluded all surgical APR-DRGs in MDC '05' ('Diseases and Disorders of the Circulatory System'). For GI Hemorrhage mortality, we excluded all medical APR-DRGs in MDC '06' ('Diseases and Disorders of the Digestive System'). Some of the indicators fall into only a few DRG categories. All discharges with carotid endarterectomy, for example, were within DRG '005', ('Extracranial Vascular Procedures'). For these indicators, we relied primarily on the severity subclass, which is independent of the DRG.

Actual implementation of the model involves running a regression with potentially a few thousand variables (each DRG divided into four severity subclasses) on millions of observations, straining the capacity of most statistical software and computer systems. In order to limit the number of covariates (DRG groups) in the model, we restricted the total number to 165 categories (DRG by severity), which was for all indicators sufficient to include 80% of discharges. We maintained all severity or risk-of-mortality subgroups for each APR-DRG included in the model in our construction of the raw adjusted estimates. The adjusted estimates of hospital performance are reported and used to compute descriptive statistics for each indicator in each year, as noted below. They are also used to construct smoothed estimates of each indicator.

The risk-adjusted estimates of hospital performance (age, gender, APR-DRG) and area performance (age, gender only) were used to construct descriptive statistics and smoothed estimates for each QI, as described in the next section.

# 2.E. Empirical Methods

## Analysis Approach

*Data sources.* The data sources used in the empirical evaluation were the 1995-97 Nationwide Inpatient Sample (NIS), which has been used for previous HCUP development, and the complete State Inpatient Data (SID) for 5 HCUP participating states (California, Florida, Illinois, New York, and Pennsylvania). The annual NIS consists of about 6,000,000 discharges and over 900 hospitals. The HCUP Nationwide Inpatient Sample contains all-payer data on hospital inpatient stays from selected States (Arizona, California, Colorado, Connecticut, Florida, Illinois, Iowa, Kansas, Maryland, Massachusetts, Missouri, New Jersey, New York, Oregon, Pennsylvania, South Carolina, Tennessee, Washington, Wisconsin, and starting in 1997, Georgia, Hawaii, and Utah). All discharges from sampled hospitals are included in the NIS database. The NIS is designed to approximate a 20-percent sample of U.S. community hospitals, defined as all nonfederal, short-term, general, and other specialty hospitals, excluding hospital units of institutions. Included among community hospitals are specialty hospitals such as obstetrics-gynecology, ear-nose-throat, short-term rehabilitation, orthopedic, and pediatric. Excluded are long-term hospitals, psychiatric hospitals, and alcoholism/chemical dependency treatment facilities. A complete description of the content of the NIS, including details of the participating States discharge abstracts, can be found on the Agency for Healthcare Research and Quality web site (www.ahrq.gov/data/hcup/hcupnis.htm). The SID sample consisted of 10,000,000 discharges and over 1300 hospitals, in over 200 metropolitan areas. Only the SID empirical results are reported because the provider level results were similar in both data sources, and the SID data was needed for the direct construction of area measures. All of the quality indicators can be constructed from the Nationwide Inpatient Sample, with two caveats: first, the area measures are based on a weighted sample of discharges, and are less precise than if complete state discharge data are used, and second, even though hospital sampling for the NIS was supposed to allow construction of a representative sample at the state level, it is possible that the Metropolitan Service Area (MSA)-level samples are not representative (i.e. biased). These limitations are not applicable when using the software on the full data from the SID to construct measures based on complete data from area hospitals.

*Reported quality indicators.* All potential indicators were assessed empirically by developing and conducting statistical tests for evaluation framework criteria of precision, bias, and construct validity. For each statistical test, we calculated up to four different estimates of indicator performance. First, the raw indicator was the simple observed value (e.g., the rate or volume) for each provider or area. Second, the adjusted indicator was based on the use of multi-variate regression to account for differences among providers in demographics and co-morbidities (defined using the 3M All-Patient Refined DRG, or APR-DRG) of patients, and among areas in demographics of the population. Third, we applied univariate smoothing techniques [86] to estimate the amount of random error relative to the true difference in performance (the 'reliability') for each indicator. Fourth, we applied new multi-variate signal extraction methods[187] by combining

information from multiple indicators over several years in order to extract more quality signal from each individual indicator than is possible with the univariate methods.

*Overview of empirical analysis.* Our approach included several stages, and generated a series of analyses on potential quality indicators that sequentially assessed some of the problems identified in our literature review. For reference, we began by constructing the "raw" or minimally adjusted indicator, based on the discharge data for each hospital and census data for each area. We then "risk-adjusted" this measure through a discharge-level regression that included controls for patient mix. The hospital-level and area-level fixed effects in these regressions are the estimates of quality indicators that are typically reported for particular hospitals and areas, and they typically reflect substantial noise. In the second stage of the analysis, we then "smoothed" these estimates using a Bayesian procedure to yield a best-guess estimate of true hospital or area performance on the indicator – the "signal" in the observed noisy measure. We did this in two ways. First, we used a univariate approach, in which the distribution of the indicator itself is used to construct the best guess. This is the smoothing or shrinkage approach most widely used in the literature on provider quality.[87-89] Second, we used a multivariate approach, in which the joint distribution of a large number of indicators (and the indicator of interest in previous time periods) is used to construct the best-guess estimate of performance. In general, the covariation among different indicators and within each indicator over time implies that we can generate much more precise estimates of true hospital or area quality using this multivariate signal extraction approach. All of our estimates of factor loadings and correlations are based on smoothed estimates, which helps to improve our ability to detect correlations, thereby addressing the multidimensionality of quality. Finally, we also report summary statistics describing the performance of the indicator in terms of the principal domains described in our literature review: precision, bias, and construct validity.

### Intuition Behind Univariate and Multi-variate Methods

An important limitation of many quality indicators is their imprecision, which complicates the reliable identification of persistent differences among providers in performance. The imprecision in quality indicators arises from two sources. The first is sampling variation, which is a particular problem for indicators based on small numbers of patients per provider (where the particular patients treated by the provider in a given year are considered a 'sample' of the entire population who might have been treated or will be treated in the near future). The amount of variation due to the particular sample of patients is often large relative to the total amount of provider level variation that is observed in any given quality indicator. A second source of imprecision arises from non-persistent factors that are not sensitive to the size of the sample; for example, a severe winter results in higher than usual rates of pneumonia mortality. Both small samples and other one-time factors that are not sensitive to sample size can add considerable volatility to quality indicators. Also, it is not the absolute amount of imprecision that matters, but rather the amount of imprecision relative to the underlying signal (i.e., true provider level variation) that dictates the reliability of any particular indicator. Even indicators based on relatively large samples with no non-persistent factors at work can be imprecise if the true level of variation among providers is negligible.

Our approach to account for the imprecision or lack of reliability is a generalization of the idea of applying a "shrinkage factor" to each provider's estimate so that less reliable estimates are shrunk toward the national average. These 'reliability-adjusted' estimates are sometimes referred to as 'smoothed' estimates (because provider performance is less volatile over time) or 'filtered' estimates (because the methods filter out the non-systematic noise, much like a radio filters our background noise to improve the radio signal). If the observed provider variation = signal variation + noise variation, then the shrinkage factor would be signal variation / (signal variation + noise variation). For example, suppose that the observed variation among providers in the in-hospital pneumonia mortality rate was a standard deviation of 10.2 percentage points, and the signal variation was a standard deviation of 5.0 percentage points. Then the shrinkage factor for this indicator is 0.240 = (0.050^2) / (0.102^2). The generalization of this approach seeks to extract additional signal using information on the relationship among multiple indicators over time.

Many of the key ideas behind the reliability-adjusted or filtered estimates are illustrated through a simple example. Suppose that one wants to evaluate a particular provider's performance based on in-hospital mortality rates among patients admitted with pneumonia, and data is available for the most recent two years. Consider the following three possible approaches: (1) use only the most recent mortality rate, (2) construct a simple average of the mortality rates from the two recent years, or (3) ignore the provider's mortality rate and assume that mortality is equal to the national average. The best choice among these three approaches depends on two important considerations: the signal-to-noise ratio in the provider's data, and how strongly correlated performance is from one year to the next. For example, suppose that the mortality rate for the provider was based on only a few patients, and one had reason to believe that mortality did not vary much across providers. Then one would be tempted to choose the last option and ignore the provider's own data because of its low reliability (e.g. low signal-to-noise ratio). This is the idea of simple shrinkage estimators, in which less reliable estimates are shrunk toward the average for all providers. Alternatively, if one had reason to believe that provider mortality changed very slowly over time, one might choose the second option in hopes that averaging the data over two years would reduce the noise in the estimates by effectively increasing the sample size in the provider. Even with large numbers of patients one might want to average over years if idiosyncratic factors (such as a bad winter) affected mortality rates from any single year. Finally, one would tend to choose the first option, and rely solely on mortality from the most recent year, if such idiosyncratic factors were unimportant, if the provider admitted a large number of patients each year, and if mortality was likely to have changed from the previous year.

Our method of creating filtered estimates formalizes the intuition from this simple example. The filtered estimates are a combination of the provider's own quality indicator, the national average, and the provider's quality indicators from past years or other patient outcomes. As suggested by the example, to form the optimal combination one must know the amount of noise and signal variance in each indicator, as well as the correlation across indicators in the noise and signal variance. We estimate the noise variance (and covariance) in a straightforward manner for each provider, based on the number of patients on which each indicator is based. To estimate the signal variance (and covariance) for each quality indicator, we subtract the noise variance from the total

variance observed in each indicator across providers (which reflects both signal and noise variance). In other words, the observed variation in quality indicators is sure to overstate the amount of actual variation across providers (because of the noise in the indicators). Therefore, we estimate the amount of true variation in performance based on how much the observed variation exceeded what would have been expected due to sampling error. Importantly, our method does not *assume* that provider performance is correlated from one year to the next (or that performance is correlated across indicators). Instead, we estimate these correlations directly from the data, and incorporate information from past years or other indicators only to the extent that these empirically estimated correlations are large.

## Smoothed Estimates of Hospital Performance

For each hospital, we observed a vector of K adjusted indicator estimates over T years from estimating the patient-level regressions (1) run separately by year for each indicator as described in the preceding section. We know that each indicator is a noisy estimate of true hospital quality; in other words, it is likely that hospitals that performed especially well or badly on the measure did so at least in part due to chance. This fact is incorporated in Bayesian methods for constructing best-guess "posterior" estimates of true provider performance based on observed performance and the within-provider noise in the measures.

In particular, let $M_j$ be the 1xTK vector of estimated indicator performance for hospital j. Then:

(2) $$M_j = \mu_j + \varepsilon_j$$

Where $\mu_j$ is a 1xTK vector of the true hospital intercepts for hospital j, and $\varepsilon_j$ is the estimation error (which has a mean zero and is uncorrelated with $\mu_j$). Note that the variance of $\varepsilon_j$ can be estimated from the patient-level regressions, since this is simply the variance of the regression estimates $M_j$. In particular, $E(\varepsilon_{jt}' \varepsilon_{jt}) = \Omega_{jt}$ and $E(\varepsilon_{jt}' \varepsilon_{js}) = 0$ for $t \neq s$, where $\Omega_{jt}$ is the covariance matrix of the intercept estimates for hospital j in year t.

We wish to create a linear combination of each hospital's observed indicators in such a way that it minimizes the mean-squared prediction error. In other words, we would like to run the following hypothetical regression:

(3) $$\mu^k_{jt} = M_j\beta^k_{jt} + v^k_{jt}$$

but cannot do this directly, since $\mu$ is unobserved and the optimal $\beta$ varies by hospital and year. While we cannot directly estimate equation (3), it is possible to estimate the parameters for this hypothetical regression. In general, the minimum mean squared error linear predictor of $\mu$ is given by $M_j \beta$, where $\beta = [E(M_j'M_j)]^{-1} E(M_j' \mu_j)$. This best linear predictor depends on two moment matrices:

(4.1) $$E(M_j'M_j) = E(\mu_j' \mu_j) + E(\varepsilon_j' \varepsilon_j)$$
(4.2) $$E(M_j'\mu_j) = E(\mu_j' \mu_j)$$

We estimate the required moment matrices directly as follows:

- Estimate $E(\varepsilon_j' \varepsilon_j)$ with the patient-level OLS estimate of the covariance matrix for the parameter estimates $M_j$. Call this estimate $S_j$. Note that $S_j$ varies across hospitals.
- Estimate $E(\mu_j' \mu_j)$ by noting that $E(M_j'M_j - S_j) = E(\mu_j' \mu_j)$. If we assume that $E(\mu_j' \mu_j)$ is the same for all hospitals, then it can be estimated by the sample average of $M_j'M_j - S_j$. Note that it is easy to relax the assumption that $E(\mu_j' \mu_j)$ is the same for all hospitals by calculating $M_j'M_j - S_j$ for subgroups of hospitals.

With estimates of $E(\mu_j' \mu_j)$ and $E(\varepsilon_j' \varepsilon_j)$, one can form least squares estimates of the parameters in equation 3 which minimize the mean squared error. Analogous to simple regression, the prediction of a hospital's true intercepts is given by:

$$(5) \qquad \hat{m}_j = M_j E(M_j' M_j)^{-1} E(M_j' m_j) = M_j [E(m_j' m_j) + E(e_j' e_j)]^{-1} E(m_j' m_j)$$

using estimates of $E(\mu_j' \mu_j)$ and $E(\varepsilon_j' \varepsilon_j)$ in place of their true values. One can use the estimated moments to calculate other statistics of interest as well, such as the standard error of the prediction and the r-squared for equation 3, based on the usual least squares formulas. We refer to estimates based on equation (5) as "filtered" estimates, since the key advantage of such estimates is that they optimally filter out the estimation error in the raw quality indicators.

Equation 5 in combination with estimates of the required moment matrices provides the basis for estimates of hospital quality. Such estimates of hospital quality have a number of attractive properties. First, they incorporate information in a systematic way from many outcome indicators and many years into the predictions of any one outcome. Moreover, if the moment matrices were known, our estimates of hospital quality represent the optimal linear predictors, based on a mean squared error criterion. Finally, these estimates maintain many of the attractive aspects of existing Bayesian approaches [87], while dramatically simplifying the complexity of the estimation. It is possible to construct univariate smoothed estimates of hospital quality, based only on empirical estimates for particular measures, using the models just described but restricting the dimension of $M_j$ to only a particular indicator k and time period t. Of course, to the extent that the provider indicators are correlated with each other and over time, this will result in a less precise (efficient) estimate.

With many years of data, it helps to impose some structure on $E(\mu_j'\mu_j)$ for two reasons. First, this improves the precision of the estimated moments by limiting the number of parameters that need to be estimated. Second, a time series structure allows for out-of-sample forecasts. We use a non-stationary, first-order Vector Autoregression structure (VAR). The VAR model is a generalization of the usual autoregressive model, and assumes that each hospital's quality indicators in a given year depend on the hospital's quality indicators in past years plus a contemporaneous shock that may be correlated across quality indicators. In most of what follows, we assume a non-stationary first-order VAR for $\mu_{jt}$ (1xK), where:

$$(6) \qquad \mu_{jt} = \mu_{j,t-1}\Phi + u_{jt}\ , \text{ with } V(u_{jt}) = \Sigma \text{ and } V(\mu_{j1}) = \Gamma\ .$$

Thus, we need estimates of the lag coefficient ($\Phi$), the variance matrix of the innovations ($\Sigma$), and the initial variance condition ($\Sigma$), where $\Sigma$ and $\Gamma$ are symmetric KxK matrices of parameters and $\Phi$ is a general KxK matrix of parameters, for a total of $2K^2+K$ parameters. For example, we must estimate 10 parameters for a VAR model with two outcomes (K=2).

The VAR structure implies that $E(M_j'M_j - S_j) = E(\mu_j'\mu_j) = f(\Phi,\Sigma,\Gamma)$. Thus, the VAR parameters can be estimated by Optimal Minimum Distance (OMD) methods,[185] i.e. by choosing the VAR parameters so that the theoretical moment matrix, $f(\Phi,\Sigma,\Gamma)$, is as close as possible to the corresponding sample moments from the sample average of $M_j'M_j - S_j$. More specifically, let $d_j$ be a vector of the non-redundant (lower triangular) elements of $M_j'M_j - S_j$, and let $\delta$ be a vector of the corresponding moments from the true moment matrix, so that $\delta=g(\Phi,\Sigma,\Gamma)$. Then the OMD estimates of $(\Phi,\Sigma,\Gamma)$ minimize the following OMD objective function:

$$(7) \qquad q = N\left[\bar{d} - g(\Phi,\Sigma,\Gamma)\right]' V^{-1}\left[\bar{d} - g(\Phi,\Sigma,\Gamma)\right]$$

where V is the sample estimate of the covariance matrix for d, and $\bar{d}$ is the sample average of d. If the VAR model is correct, the value of the objective function, q, will be distributed $\chi^2$ (p) where p is the degree of over-identification (the difference between the number of elements in d and the number of parameters being estimated). Thus, q provides a goodness of fit statistic that indicates how well the VAR model fits the actual covariances in the data.

Finally, we use estimated $R^2$ statistics to evaluate the filtered estimates' ability to predict (in sample) and forecast (out-of-sample) variation in the true intercepts, and to compare our methods to conventional methods (e.g. simple averages, or univariate shrinkage estimators). If true hospital intercepts ($\mu$) were observed, a natural metric for evaluating the predictions would be the sample r- squared:

$$(8) \qquad R^2 = 1 - \left(\sum_{j=1}^{N} \hat{u}_j^2\right) \Big/ \left(\sum_{j=1}^{N} m_j^2\right),$$

where $\hat{u}_j = m_j - \hat{m}_j$ is the prediction error. Of course $\mu$ is not observed. Therefore, we construct an estimate using our estimate of $E(\mu_j' \mu_j)$ for the denominator, and our estimate of $E\left[\left(m_j - \bar{m}_j\right)'\left(m_j - \bar{m}_j\right)\right]$ for the terms in the numerator (where this can be constructed from the estimated moment matrices in equations 4.1-4.2). Finally, we report a weighted R-squared (weighting by the number of patients treated by each hospital).

As in earlier work using this method for cardiac care in the adult population, we validate our indicators using out-of-sample performance, based on forecasts (e.g., using the first two years of data to predict in subsequent year) and based on split-sample prediction (e.g., using one half of the patient sample to predict outcome indicators in the other half of the sample). For evaluating out-of-sample forecasts, we construct a modified R-squared of the forecast that estimates the fraction of the *systematic* (true) hospital variation in the outcome measure (M) that was explained:

(9) $$\tilde{R}^2 = 1 - \left( \sum_{j=1}^{N} \left( \hat{v}_j^2 - S_j \right) \right) \Big/ \left( \sum_{j=1}^{N} \left( M_j^2 - S_j \right) \right),$$

where $\hat{v}_j = M_j - \hat{m}_j$ is the forecast error, and $S_j$ is the OLS estimate of the variance of the estimate $M_j$. This modified R-squared estimates the amount of variance in the true hospital effects that has been forecasted. Note that because these are out-of-sample forecasts, the R-squared can be negative indicating that the forecast performed worse than a naive forecast in which one assumed that quality was equal to the national average at all hospitals.

## Empirical Analysis Statistics

Using the methods just described, we constructed a set of statistical tests to evaluate precision, bias, and construct validity. Each of the key statistical test results for these evaluation criteria was summarized and explained in the overview section of this chapter. Tables 1-3 provides a summary of the statistical analyses and their interpretation. Indicators were tested for precision first, and ones that performed poorly were eliminated from further consideration. These indicators are noted in Appendix 5. We assessed bias and construct validity for all recommended indicators. Cut-offs and further interpretation details are provided with results tables.

**Table 1. Precision Tests**

| Measure | Statistic | | Interpretation |
|---|---|---|---|
| Precision. Is most of the variation in an indicator at the level of the provider? Do smoothed estimates of quality lead to more precise measures? | | | |
| a. Raw variation in indicator | ▪ Provider Standard Deviation<br>▪ Signal Standard Deviation<br>▪ Provider/Area Share | - Unadjusted<br>- Age-sex adjusted<br>- Age-sex + APR-DRG adjusted | Provider variation is signal variation + noise variation. What percentage of the total variation (patient + provider) is between-provider variation (a measure of how much variation is subject to provider control). Risk adjustment can either increase or decrease true variation. |
| b. Univariate smoothing | • Signal/(Signal-to-Noise) Ratio:<br>- Unadjusted<br>- Age-sex adjusted<br>- Age-sex + APR-DRG adjusted | | Estimates what percentage of the observed variation between providers reflects 'true' quality differences versus random noise. Risk adjustment can increase or decrease estimates of 'true' quality differences. |
| c. MSX methods | • In-sample R-squared:<br>- Unadjusted<br>- Age-sex adjusted<br>- Age-sex + APR-DRG adjusted | | To the extent that indicators are correlated with each other and over time, MSX methods can extract more 'signal' (a higher percentage of observed variation between providers that reflects 'true' quality). |

**Table 2. Bias Tests**

| Measure | Statistic | Interpretation |
|---|---|---|
| Bias. Does risk-adjustment change our assessment of relative provider performance, after accounting for reliability? Is the impact greatest among the best or worst performers, or overall? What is the magnitude of the change in performance? | | |
| a. MSX methods: unadjusted vs. age, sex, APR-DRG risk-adjustment | • Rank Correlation Coefficient (Spearman) | Risk-adjustment matters to the extent that it alters the assessment of relative provider performance. This test determines the impact overall. |
| | • Average Absolute Value Of Change Relative To Mean | This test determines whether the absolute change in performance was large or small relative to the overall mean. |
| | • Percentage of The Top 10% Of Providers That Remains The Same | This test measures the impact at the highest rates (in general, the worse performers, except for measures like VBAC). |
| | • Percentage of The Bottom 10% Of Providers That Remains The Same | This test measures the impact at the lowest rates (in general, the best performers, except for measures like VBAC). |
| | • Percentage of providers that move more than two deciles in rank (up or down) | This test determines the magnitude of the relative changes. |

**Table 3. Construct Validity Tests**

| Measure | Statistic | Interpretation |
|---|---|---|
| 3. Construct Validity. Is the indicator related to other indicators in a way that makes clinical sense? Do methods that remove noise and bias make the relationship clearer? | | |
| a. Correlation of indicator with other indicators | Pearson correlation coefficient | Are indicators correlated with other indicators in the direction one might expect? |
| b. Factor loadings of indicator with other indicators | Factor loadings | Do indicators load on factors with other indicators that one might expect? |

# Chapter 3: Results

The results chapter is divided into 5 sections. *Each indicator is assigned an identification number used throughout the results section and in the appendices.* The sections are designed to offer the results in a comprehensive and clear manner. Each section is designed to present the results in a unique way, and thus some redundancy is inevitable.

**1. Summary of Evidence for Indicators (Section 3.A.)** presents the literature review findings and results of the empirical evaluation for the final indicators that comprise the HCUP II QIs. This section is intended for the reader who does not wish to delve into the details of the evaluation (which are presented in section 5). There are 4 main subsections:

> ***Initial Empirical Evaluation*** (Section 3.A.1.) provides a description of the winnowing of over 200 indicators down to 45 recommended indicators.
> ***Summary of Literature Review and Empirical Evaluation*** (Section 3.A.2) provides a text summary of the literature review and empirical analyses findings for provider-level and then area-level indicators. The text summaries refer to the tables contained in the following two sub-sections.
> ***Summary Tables by Indicator*** (Section 3.A.3, Tables 9-13) provides supporting evidence from the literature review and empirical analysis along with suggestions for use, by indicator, organized by indicator type.
> ***Summary Tables of Evidence by Empirical Test*** (Section 3.A.4, Tables 14-25) provides a succinct synopsis of all empirical findings. The results are organized by the tests for the volume indicators, and of precision, minimum bias, and construct validity.

Sections 2-4 present results for specific aspects of the review and evaluation:
**2. Results of Semi-Structured Interviews (Section 3.B.)** provides information from a variety of organizations on the application of indicators and risk adjustment methods as well as practical suggestions for refinement of the HCUP I QIs.
**3. Review of Risk Adjustment Approaches (Section 3.C.)** provides the rationale for the use of APR-DRGs as a risk adjustment approach for this version of the QIs.
**4. Evidence from Literature by Indicator Type (Section 3.D.)** provides a broad overview of quality indicators in general. This is a literature review of the use and validity for each of the major subgroups of indicators: volume, utilization, ambulatory sensitive conditions, and mortality. Because indicators within a subgroup often have similar limitations and because subgroups of indicators, instead of specific individual indicators, are often reported in the literature, this section offers a general overview of the literature findings.

The final section provides the details of the review, organized by indicator:
**5. Detailed Evidence by Indicator (Section 3.E)** provides a comprehensive presentation of the literature review and the empirical evaluation for each indicator that is included in the HCUP II QIs. This section consists of detailed textual summaries of the evidence for each indicator. Each indicator summary begins with a definition of the indicator evaluated, followed by the findings of the literature review and the empirical analyses. At the end of each write-up, a discussion paragraph integrates the empirical and literature findings and provides recommendations for use.

# 3. A. Summary of Evidence for Indicators

This section summarizes the results on the recommended quality indicators, including both the literature review and the empirical analyses. These results are presented in four sections, described in the introduction to this chapter.

## 3.A.1. Initial Empirical Evaluation

As described in Section 2.C. "Literature Review Methods" over 200 indicators were identified from the literature, databases of indicators, and through personal contact. These indicators are listed in Appendix 5, and represent indicators definable using administrative data, and applicable to a large share of providers or areas (i.e. does not include highly specialized clinical areas, such as burn-units). Seventy-one indicators were selected for initial empirical evaluation, according to the criteria outlined in Section 2.C., including all non-complication HCUP I QIs. In general these indicators related to a relatively large number of patients and or hospitals, had adequate face validity, and provided a comprehensive examination of multiple aspects of health care. Reasons for selection and exclusion are also listed in Appendix 5. The initial empirical evaluation evaluated 71 indicators using tests of precision. Those indicators with low precision (less than 1.5% provider variation share or 0.01% for area variation share) were excluded from further evaluation, since the interpretation of indicators without adequate precision is unclear. Table 4 lists the 23 non-complications related HCUP I QIs tested, and their inclusion status in HCUP II. Table 5 shows the additional non-HCUP I indicators removed from further consideration exclusively because of the precision test. Several other indicators initially tested were length of stay measures, and were excluded based on input from experts regarding reservations about using them for quality assessments.

**Table 4. List of HCUP I indicators, and inclusion status in HCUP II.**

| Indicator type | Indicator name | Included in HCUP II |
|---|---|---|
| ACSC | Low birth weight | * |
| | Very low birth weight | |
| | Pediatric asthma discharges | * |
| | Immunization preventable pneumonia (combined with bacterial pneumonia after excluded for precision) | |
| | Cerebrovascular disease in the non-elderly | |
| | Diabetes – short term complications | * |
| | Diabetes – long term complications | * |
| | Perforated appendix | * |
| Utilization | Radical prostatectomy | |
| | Hysterectomy | * |
| | Laminectomy and/or spinal fusion | * |
| | TURP | |
| | CABG | *[a] |
| | Incidental appendectomy | * |
| | Cesarean section delivery | * |
| | Successful VBAC | * |
| | Laparoscopic cholecystectomy | * |
| Mortality, post-procedural | Knee replacement | |
| | Laminectomy | |
| | TURP | |
| | Cholecystectomy | |
| | Hysterectomy | |
| | Hip replacement | * |

[a]Not recommended as a stand-alone indicator.

**Table 5. List of other indicators, tested but not recommended due to low precision**

| Indicator type | Indicator name |
|---|---|
| ACSC, area-level | Gastroenteritis |
| | Severe ENT infection |
| | Cellulitis |
| | Convulsion |
| | Grand Mal |
| | Skin graft with cellulitis |
| | Hypoglycemia |
| | Pediatric diabetes |
| Utilization, area-level | Radical prostatectomy over 75 years |
| | Carotid endarterectomy |
| Mortality, medical | Heart failure and shock[a] |
| Mortality, post-procedural | Carotid endarterectomy |
| | PTCA |
| Other | Domestic violence |

[a]Definition differs from the recommended Congestive Heart Failure Mortality Indicator (#33)

## 3.A.2. Summary of Results

Our evidence report recommends 25 provider-level quality indicators and 20 area-level indicators for use (see Tables 6 and 7) (See Appendix 6 for full definitions of recommended indicators). Provider-level quality indicators measure hospital or other provider quality, and are defined with a provider-level denominator. Area-level quality indicators most likely measure health system quality within an area and are defined with a population denominator. While none of these indicators is without its limitations, a considerable literature in most cases coupled with evidence of satisfactory empirical performance suggests that the recommended indicators may be useful additions to the "toolkit" of a broad range of clinical quality improvement professionals, health care managers, health policymakers, as well as researchers. Each of these indicators, when used with the appropriate caveats in mind, is appropriate for screening for quality problems – as a first step in identifying potential quality problems. The accompanying Tables review the results of the detailed evaluations of all the indicators. The overall indicator summary tables (Tables 9, 10, 13 for the provider indicators and Tables 11-12 for the area indicators) include major findings from the existing research literature on each indicator, as well as a consistent set of empirical evaluations of indicator performance. These empirical analyses can be replicated or extended using the software that accompanies this report. Based on the literature review and empirical evaluation, the evidence summary tables (Tables 9-13) also outlines some specific guidance for using each indicator as part of a program to improve quality.

**Table 6. Provider indicator list**

| Provider-Level Indicator Name | Type of Indicator | Indicator Description | Risk Adjustment Recommended | Better quality may be associated with: |
|---|---|---|---|---|
| AAA repair volume (#1) | Volume | Raw volume compared to annual thresholds (20, 32 procedures) | Not applicable | Higher rates |
| Carotid endarterectomy volume (#2) | Volume | Raw volume compared to annual thresholds (50, 101 procedures) | Not applicable | Higher rates |
| CABG volume (#3) | Volume | Raw volume compared to annual thresholds (100, 200 procedures) | Not applicable | Higher rates |
| Esophageal resection volume (#4) | Volume | Raw volume compared to annual thresholds (6, 7 procedures) | Not applicable | Higher rates |
| Pancreatic resection volume (#5) | Volume | Raw volume compared to annual thresholds (10, 11 procedures) | Not applicable | Higher rates |
| Pediatric heart surgery volume (#6) | Volume | Raw volume compared to annual thresholds (100 procedures) | Not applicable | Higher rates |
| PTCA volume (#7) | Volume | Raw volume compared to annual thresholds (200, 400 procedures) | Not applicable | Higher rates |
| Cesarean section rate (#8) | Utilization | Number of cesarean sections per 100 deliveries | Age adjustment, and potentially supplemental (clinical data, linked to infant record, or linked to birth record). | Lower rates[b] |
| Incidental appendectomy among elderly rate (#9) | Utilization | Number of incidental appendectomies per 100 abdominal surgeries | None | Lower rates |
| Bi-lateral cardiac catheterization rate (#10) | Utilization | Number of bilateral catheterizations per 100 cardiac catheterizations | None | Lower rates |
| VBAC rate (#11) | Utilization | Number of vaginal births per 100 deliveries in women with previous cesarean section | Age adjustment, and potentially supplemental (clinical data, linked to infant record, or linked to birth record). | Higher rates[b] |
| Laparoscopic cholecystectomy (#12) | Utilization | Number of lap. cholecystectomies per 100 cholecystectomies. | Age and sex adjustment, and potentially supplemental clinical. | Higher rates[b] |
| AMI mortality (#33) | In-hospital Mortality | Number of deaths per 100 discharges for AMI | APR-DRG | Lower rates |
| CHF mortality (#34) | In-hospital Mortality | Number of deaths per 100 discharges for CHF | APR-DRG | Lower rates |

Section 3.A. Summary of Results

| Provider-Level Indicator Name | Type of Indicator | Indicator Description | Risk Adjustment Recommended | Better quality may be associated with: |
|---|---|---|---|---|
| GI hemorrhage mortality (#35) | In-hospital Mortality | Number of deaths per 100 discharges for GI hemorrhage | APR-DRG | Lower rates |
| Hip fracture mortality (#36) | In-hospital Mortality | Number of deaths per 100 discharges for hip fracture | APR-DRG | Lower rates |
| Pneumonia mortality (#37) | In-hospital Mortality | Number of deaths per 100 discharges for pneumonia | APR-DRG | Lower rates |
| Stroke mortality (#38) | In-hospital Mortality | Number of deaths per 100 discharges for stroke | APR-DRG | Lower rates |
| AAA repair mortality (#39) | Post-procedural Mortality | Number of deaths per 100 AAA repairs | APR-DRG | Lower rates |
| CABG mortality (#40) | Post-procedural Mortality | Number of deaths per 100 CABG procedures | APR-DRG | Lower rates |
| Craniotomy mortality (#41) | Post-procedural Mortality | Number of deaths per 100 craniotomies | APR-DRG | Lower rates |
| Esophageal resection mortality (#42) | Post-procedural Mortality | Number of deaths per 100 esophageal resections for cancer | APR-DRG | Lower rates |
| Hip replacement mortality (#43) | Post-procedural Mortality | Number of deaths per 100 hip replacements | APR-DRG | Lower rates |
| Pancreatic resection mortality (#44) | Post-procedural Mortality | Number of deaths per 100 pancreatic resections for cancer | APR-DRG | Lower rates |
| Pediatric heart surgery mortality (#45)[a] | Post-procedural Mortality | Number of deaths per 100 heart surgeries in patients under age 18 years | APR-DRG and potentially additional clinical data | Lower rates |

[a]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.
[b]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.

**Table 7. Area indicator list**

| Area-Level Indicator Name | Type of Indicator | Indicator Description | Risk Adjustment Recommended | Better quality may be associated with: |
|---|---|---|---|---|
| CABG rate (#13)[a] | Utilization | Number of CABGs per 100,000 population | Age and sex | Lower rates[b] |
| Hysterectomy rate (#14) | Utilization | Number of hysterectomies per 100,000 population | Age and possibly additional such as parity | Lower rates[b] |
| Laminectomy rate (#15) | Utilization | Number of laminectomies per 100,000 population | Age and sex | Lower rates[b] |
| PTCA rate (#16)[a] | Utilization | Number of PTCAs per 100,000 population | Age and sex | Lower rates[b] |
| Dehydration admission rate (#17)[a] | ACSC | Number of admissions for dehydration per 100,000 population | Age and sex | Lower rates[b] |
| Bacterial pneumonia admission rate (#18)[a] | ACSC | Number of admissions for bacterial pneumonia per 100,000 population | Age and sex | Lower rates[b] |
| Urinary infection admission rate (#19)[a] | ACSC | Number of admissions for urinary infection per 100,000 population | Age and sex | Lower rates[b] |
| Perforated appendix admission rate (#20)[a] | ACSC | Number of admissions for perforated per 100 admissions for appendicitis within an area | Age and sex | Lower rates |
| Angina admission rate (#21)[a] | ACSC | Number of admissions for angina per 100,000 population | Age and sex | Lower rates[b] |
| Asthma (adult) admission rate (#22) | ACSC | Number of admissions for adult asthma per 100,000 population | Age and sex | |
| COPD admission rate (#23) | ACSC | Number of admissions for COPD per 100,000 population | Age and sex. Potential supplement includes patient characteristics, such as smoking. | Lower rates[b] |
| CHF admission rate (#24) | ACSC | Number of admissions for CHF per 100,000 population | Age and sex. | Lower rates[b] |
| Diabetes short term complication admission rate (#25) | ACSC | Number of admissions for diabetes short term complications per 100,000 population | Age and sex. Potential supplement includes population diabetes incidence rates. | Lower rates |
| Uncontrolled diabetes admission rate (#26)[a] | ACSC | Number of admissions for uncontrolled diabetes per 100,000 population. | Age and sex. Potential supplement includes population diabetes incidence rates. | Lower rates[b] |
| Diabetes long term | ACSC | Number of admissions for | Age and sex. Potential | Lower rates |

| Area-Level Indicator Name | Type of Indicator | Indicator Description | Risk Adjustment Recommended | Better quality may be associated with: |
|---|---|---|---|---|
| complication admission rate (#27) | | diabetes long term diabetes per 100,000 population | supplement includes population diabetes incidence rates. | |
| Lower extremity amputation in diabetics rate (#28)[a] | ACSC | Number of LE amputations in diabetics per 100,000 population. | Age and sex | Lower rates |
| Hypertension admission rate (#29) | ACSC | Number of admissions for hypertension per 100,000 population | Age and sex. Potential supplement includes population diabetes incidence rates. | Lower rates[b] |
| Low birth weight rate (#30)[a] | ACSC | Number low birth weight births per 100 births in an area | None available in HCUP. Potential supplement include clinical, link to mother's record, or link to birth record. | Lower rates |
| Pediatric asthma admission rate (#31) | ACSC | Number of admissions for pediatric asthma per 100,000 population | Age and sex | Lower rates[b] |
| Pediatric gastroenteritis admission rate (#32)[a] | ACSC | Number of admissions for pediatric gastroenteritis per 100,000 population | Age and sex | Lower rates[b] |

[a] 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.
[b] 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.

## Provider Indicators

*Provider indicators* are constructed at the hospital level; they provide information related to the quality of care at individual hospitals. There are several types of indicators included:

- *Volume indicators* for inpatient procedures where substantial evidence of an important volume-outcome relationship has been demonstrated. These indicators include abdominal aortic aneurysm repair volume, carotid endarterectomy volume, CABG volume, esophageal resection volume, pancreatic resection volume, pediatric heart surgery volume, and PTCA volume.

- *Utilization indicators* for 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 or higher costs or both. These indicators include cesarean section delivery rate, incidental appendectomy in the elderly rate, bi-lateral catheterization rate, successful vaginal birth after cesarean section rate, and laparoscopic cholecystectomy rate.

- *Mortality indicators for inpatient procedures* for which mortality rates have been shown to vary substantially across institutions and for which evidence suggests that high mortality, at least in part, may be associated with deficiencies in the quality of care. These indicators include mortality for acute myocardial infarction, congestive heart failure, gastrointestinal hemorrhage, hip fracture, and stroke.

- *Mortality indicators for inpatient conditions* for which mortality rates have also been shown to vary substantially across institutions, and for which evidence suggests that high mortality, at least in part, may be associated with deficiencies in the quality of care. These indicators include mortality after abdominal aortic aneurysm repair, CABG, craniotomy, esophageal resection, hip replacement, pancreatic resection, and pediatric heart surgery.

By level of evidence, Tables 9 (volume), 10 (utilization) and 13 (mortality) summarize our literature review and empirical evaluation of each of these indicators. While each provider-level indicator features distinctive issues, a number of common themes are apparent across many of these indicator types. We review these issues in the following subsections.

### Volume Indicators: Reliably-Measured Quality "Proxies"

The volume indicators are somewhat different from the other provider-level indicators, in that they simply represent counts of admissions in which particular major procedures were performed, rather than more direct measures of performance. As such, they are not subject to many of the issues of noise and bias that interfere with the interpretation of other provider-level quality indicators, as discussed in the next subsection. The recommended volume indicators include those for which substantial research has demonstrated 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 weakest evidence linking volume and outcome exists for coronary artery bypass surgery, for which recent studies that

included clinically detailed risk adjusters found a weak and statistically insignificant relationship. We retained this indicator in our recommended set because of its historical use, because these recent studies came from areas with few low-volume hospitals, and because surgeon volume still appears to be a significant predictor of mortality. For the other volume indicators, the empirical evidence of a relationship is considerably stronger.

Administrative data, like the HCUP data and the state hospital discharge data from which they are derived, may be the best available source for accurate and comprehensive counts of major inpatient procedures performed by hospitals. For all of the recommended volume indicators, there is little evidence that procedures are miscoded or not reported. For one of the recommended volume indicators (PTCA), a small fraction of procedures (10% or less) are performed on an outpatient basis, and hence are omitted from HCUP, but these missing count data should only rarely influence conclusions about whether providers exceed recommended volume thresholds. For two of the recommended volume indicators (esophageal resection and pancreatic resection in cancer patients), the procedures were so infrequent that counts from a single year may not provide reliable measures of hospital volume and experience. However, counts over several years can provide a quite reliable measure of hospital volume, even in these cases. Thus, the HCUP II volume indicators can provide valuable information for health policymakers, purchasers, consumers, and others on whether particular hospitals meet recommended volume thresholds.

Tables 14-17, based on results from the analysis of 1995-97 HCUP data that are presented in detail in the report Appendix 7, summarizes the distribution of procedure volumes for the recommended indicators among HCUP hospitals. The Table shows that the vast majority of adult patients undergoing cardiac procedures – PTCA and CABG – were treated by hospitals that meet at least a "lower" recommended volume threshold. However, patients undergoing carotid endarterectomy and abdominal aortic aneurysm repair were considerably less likely to be treated by hospitals that met recommended volume thresholds. The relatively few patients undergoing esophageal or pancreatic resection were unlikely to be treated by hospitals that met recommended volume thresholds. Finally, a significant fraction of pediatric heart surgery patients were treated by hospitals that did not meet the volume thresholds. However, as we note in our review of this indicator, pediatric heart surgery consists of a heterogeneous set of procedures; thus, it is possible that hospitals with low overall surgery volume had relatively high volumes for particular types of specialized surgery. None of these results showed substantial differences between 1995 and 1997.

Table 18 summarizes the correlations among the hospital volume measures, and the correlations between volume measures and the associated hospital procedure mortality measures (described in the subsection, Mortality Indicators). Not surprisingly, hospital volumes for CABG, PTCA, carotid endarterectomy, and abdominal aortic aneurysm repair are strongly correlated. The weaker correlations between the volumes of these procedures and pediatric heart surgery volume reflects the fact that some pediatric surgery centers tend to specialize in pediatric care, and so have lower or no volume of adult surgical procedures. The hospitals specializing in cancer surgery also tend to be somewhat different from those performing substantial cardiovascular surgery. In general, higher-volume hospitals tend to have lower inpatient mortality rates for that procedure, as well as lower inpatient mortality for related procedures.

These empirical results confirm that hospital volume is an important correlate of quality of care. However, these results as well as the other analyses and prior studies summarized in our detailed review of each indicator also demonstrate that volume is at best a quite noisy reflection

of quality. 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 weak relationship between volume and quality include the possible importance of surgeon volume (not captured presently in HCUP data), differences in the severity and complexity of cases treated, and many other factors. Moreover, use of volume as a quality indicator may lead to undesirable hospital responses, such as the performance of more procedures on patients with 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 quality, to help determine whether a high-volume hospital actually provides high-quality care, and whether a low-volume hospital provides low-quality care.

**Utilization Indicators (provider-level only)**

*Precision.* All of the recommended indicators show a large amount of variation across hospitals, suggesting that important opportunities for improving quality of care exist. In addition, with the exception of incidental appendectomy, all of these indicators involve common procedures (i.e., large numerators and denominators), and so they are all, at least, relatively precisely measured. "Smoothing" and related methods to account for random noise are helpful, but except for incidental appendectomy, such methods have a relatively modest impact on the measured performance of medium-to-high volume hospitals; Methods to account for random noise are less critical for the utilization measures than for the hospital mortality measures.

*Minimum bias.* While risk adjustment (where feasible) has some impact on measured performance for some of the measures, none of the measures appear to be highly sensitive to risk adjustment based on age, sex, and APR-DRGs (where applicable). Nevertheless, because differences in patient characteristics that are not captured through such risk-adjustment may influence whether or not the procedure is appropriate, the utilization measures may still be biased. In addition, all cesarean deliveries are classified to APR-DRG 540, and all cholecystectomies are classified as either laparoscopic (APR-DRG 263) or non-laparoscopic (APR-DRG 262). APR-DRGs cannot be used to adjust for patients' underlying risk of cesarean delivery, vaginal birth after cesarean delivery, or laparoscopic cholecystectomy, because they are assigned based on utilization of the procedure. More careful risk-adjustment, even based on the limited clinical information available from HCUP data, could identify and remove additional bias.

*Construct validity.* Though the ideal rate for each of the indicators has not been established, substantial evidence suggests that the rates observed at many hospitals are very likely to be inappropriate. This is particularly true for bilateral catheterization and incidental appendectomy, which have very low optimal rates at virtually all hospitals, although both of these procedures are still commonly performed. Multiple studies suggest that high-quality centers can safely reduce the utilization of these procedures to relatively low rates. In addition, empirical analysis suggests that several of these measures may be correlated with each other. For example, hospitals with higher cesarean delivery rates tend to have lower rates of vaginal birth after cesarean and higher rates of incidental appendectomy. However, it is not clear whether utilization indicators are useful "proxy" indicators for other aspects of hospital quality.

*Fosters true quality improvement.* In summary, compared to the mortality indicators, the utilization indicators require somewhat less sophisticated statistical methods and provide relatively clear evidence of likely quality differences across hospitals. Although there are very few indications for incidental appendectomy and bilateral catheterization, the "right" utilization rate for these indicators is generally not known. Thus, it may be more useful to identify very high or low outliers than to emphasize numerical rate differences. In addition, for the elective non-obstetrical procedures, use of these quality indicators may create a perverse incentive to increase the denominator for the utilization measures. For example, hospitals could increase their laparoscopic cholecystectomy rate by performing more cholecystectomies on low-risk, low-benefit patients, rather than by shifting patients from open to laparoscopic surgery. Simultaneous evaluation of the area rate of the denominator procedures (i.e., any type of cholecystectomy) can provide some evidence on this question. Finally, use of these indicators could induce under-reporting.

*Further investigation.* These indicators are likely to be most useful as a "screen" for further evaluations, using supplemental data to determine whether utilization is truly inappropriate. Incidental appendectomy is generally inappropriate, but review of a few of the cases performed might identify valid exceptions. For all of the remaining procedures, 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 a significant number of inappropriate cases.

**Mortality Indicators**

HCUP data can be used to construct a number of indicators for inpatient mortality after major procedures and for common medical conditions leading to hospitalization. Because patient characteristics beyond the control of medical providers are the primary determinants of mortality, and because mortality is a relatively infrequent outcome for most conditions, the problems of noise and bias are more substantial concerns with the mortality indicators than with the other recommended hospital quality indicators. On the other hand, the recommended mortality indicators demonstrate generally large differences across hospitals that do not appear to be due to random chance or to differences in comorbid diseases reported in hospital discharge data. Because mortality is a very important outcome, these indicators can potentially be used as part of a careful and thorough quality improvement effort to reduce inpatient mortality.

*Precision.* For most mortality indicators, a substantial part of the apparent variation in mortality rates across hospitals can be attributed to unsystematic unobservable characteristics, or noise. Some mortality indicators (e.g., hip fracture) have relatively less noise than others (e.g., hip replacement and GI hemorrhage). Three indicators, AAA repair, esophageal resection, and pancreatic resection appear particularly noisy. Nonetheless for all of the mortality indicators, except hip replacement, methods to account for unsystematic variation in developing a "best estimate" of hospital performance, such as the smoothing methods applied in this report, have a substantial impact on the indicators and are strongly recommended for all mortality indicators to avoid misidentifying "outlier" providers and to develop a reasonably good forecast of a provider's mortality rate in the future.

*Minimum bias.* Because patient characteristics are relatively important determinants of mortality, these measures also have a large potential for bias, due to unobserved differences in case mix. Biases may also arise in all of the measures because of differences in hospital discharge and transfer practices, because HCUP currently does not permit information on deaths after discharge or certain transfers to be included. As Table 13 and our associated literature reviews note, several studies have provided evidence on the importance of such biases for most of the recommended indicators. In most cases, risk adjustment using secondary diagnosis information in HCUP data can help address the problem of case-mix differences. Thus, such risk adjustment should be performed when feasible. For a few indicators, evidence suggests that the particular choice of risk adjustment method (e.g., APR-DRG or Comprehensive Severity Index (CSI), as described in more detail in section 2.D. Risk Adjustment Methods) may affect rankings. However, these previous studies of the effects of alternative risk adjustment systems generally did not account for random variation in hospital level performance. As a result, quite noisy measures of hospital mortality were compared, so it is perhaps not surprising that the results varied by risk adjustment methodology. HCUP users should look for future studies that compare the effects of different risk adjustment systems on *smoothed* hospital mortality indicators. In the meantime, however, users should realize that measured performance for some indicators might depend on the risk adjustment method that is applied.

Another potential source of bias for some of the mortality indicators is differences in hospital outpatient (emergency room) treatment quality for particular conditions. Better hospitals may be more capable of treating milder cases on an outpatient basis, so that the patients actually admitted with the condition have relatively high severity and comorbidity. Whereas all of the procedures included as mortality quality indicators are generally performed on an inpatient basis , many patients with relatively uncomplicated GI hemorrhage, congestive heart failure, and pneumonia, or possibly stroke may be managed effectively as outpatients in some centers.

*Construct validity.* All of these considerations suggest that hospital mortality indicators can be used most effectively as quality screens, rather than as fully validated, reliable measures of hospital performance. To the extent possible, they should be considered in conjunction with other sources of information on hospital quality, to identify opportunities for quality improvement. Some of the other sources of information can be derived from HCUP data: mortality indicators can be considered in conjunction with hospital volume indicators and patient safety indicators (e.g., long hospital stays) for the same condition or procedure, as well as mortality and other quality indicators for related procedures (e.g., after accounting for random noise, CABG and AAA mortality rates tend to be correlated with each other). A consistent pattern across many of these indicators could provide more evidence on a potential underlying quality problem. Although some hospitals excel in specific areas of care and not others, HCUP data can be used to identify relationships among quality indicators, when they exist.

*Fosters true quality improvement.* One potential concern about the use of mortality indicators is that they may provide incentives to avoid cases that are more difficult and to treat patients with milder illnesses. Although little direct evidence exists on this question, some surveys have suggested, following published reports on CABG mortality rates in Pennsylvania and New York, that it is a potential problem. However, few studies have investigated whether

improvements in quality of care resulting from use of an indicator outweigh these potential adverse consequences.

An additional concern is that use of in-hospital mortality indicators without accompanying information on post-hospitalization mortality may lead to the premature discharge of patients to die at elsewhere. While it has been shown for a few of the mortality indicators that 30-day mortality differs from in-hospital mortality, there is no evidence that use of in-hospital mortality indicators has actually created this perverse incentive.

*Further investigation.* Because of the important limitations of mortality measures based on discharge abstracts, these measures can benefit significantly from use in conjunction with other sources of data on hospital quality. One potential source of additional information is medical chart review or linkage with hospital-based clinical data systems (e.g., laboratory test results), which would allow better adjustment for severity and comorbidity. Even if a comprehensive clinical risk adjustment program is not feasible, such patient record data can be used in a limited way to determine whether the average case mix of patients treated by different hospitals differs in a manner that could explain performance on HCUP II measures. Record reviews may also be helpful for determining weaknesses in processes of care that are associated with lower mortality. Finally, because many of the indicators are significantly related to each other, 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 in many dimensions. Biases related to potentially incomplete patient follow-up can be addressed through record linkages. Better information on post-hospitalization complications 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 index (mortality) records. Finally, analyses of hospital outpatient data (particularly ambulatory surgery and emergency room data) in conjunction with inpatient data can help to determine whether variations in risk-adjusted mortality reflect differences in outpatient practices. The literature includes examples of all of these approaches.

**Summary of Empirical Evaluation of Precision and Bias for Provider indicators**

Tables 19 and 21 summarize the empirical evaluations of precision (including variation across hospitals) and bias for the hospital utilization and mortality indicators. Empirical findings on the performance of specific indicators are presented in detail in Appendix 7 and are summarized in Tables 19 and 21.

## Area Indicators

Our evidence report includes a set of quality indicators constructed at the *area* level. Area level indicators are constructed with a population denominator. Two types of indicators are included:

- *Utilization rate indicators* for procedures whose use has been shown to vary widely across relatively similar geographic areas, and which have also been shown to include substantial inappropriate and/or equivocal utilization. These recommended utilization indicators include hysterectomy and laminectomy. Two other utilization indicators (rates of coronary artery

bypass surgery and PTCA) were also included so that users may identify differences and track changes in utilization that correlate with HCUP II volume and mortality indicators.

- *Ambulatory care sensitive condition (ACSC) indicators* involving admissions for diagnoses that could have been prevented or ameliorated with currently recommended outpatient care, according to recent evidence from population-based studies. The recommended ACSC indicators include dehydration, bacterial pneumonia, urinary infection, perforated appendix, angina, adult asthma, chronic obstructive pulmonary disease, congestive heart failure, diabetes (short and long term complications, uncontrolled diabetes, and lower extremity amputation), hypertension, low birth weight, pediatric asthma and pediatric gastroenteritis.

Versions of some of these indicators were previously recommended as HCUP I QIs. However, their construction differs in HCUP II, in that the denominator for these indicators is constructed at the area level. For most of the indicators, the denominator is the age- and gender-adjusted *population* rate of hospitalization with the procedure or diagnosis. (There are two exceptions: for perforated appendix and low birth weight rate, we use as denominators all hospitalized cases of appendicitis and all births, respectively. In these cases, the indicators are constructed at the area level with denominators consisting of age- and gender- standardized rates for the population of appendectomies in the area or sex-standardized rates for the population of births in the area.) In the previous version of HCUP (HCUP I), the denominator was some set of discharges or all discharges at each *hospital*. The hospital-based indicators were criticized by many reviewers as misleading, since the hospital is not the best unit of analysis for measures that relate primarily to area health and health care. For example, a high-volume CABG hospital might have a very high CABG rate relative to other hospitals, but it could achieve very good outcomes (because of its specialization in CABG care) and actual area CABG rates might be low (because the hospital is a regional referral center). Thus, the indicator might suggest a quality problem when, in fact, none exists. By constructing ambulatory-care sensitive condition indicators at the area level, outliers for these measures will not simply be hospitals that specialize in procedures or that happen to care for a disproportionate share of patients receiving poor outpatient care.

Because HCUP data do not include specific information on patient residence (e.g., zip code), it is not possible to construct meaningful measures of area rates for very small areas. Nor is it possible to construct measures for the hospital referral regions used in the Dartmouth Atlas of Medical Care for fee-for-service Medicare patients, which also depend on information on zip code of residence. Rather, the smallest feasible area for analysis is the level that provides relatively modest "leakage" into or out of hospitals within the area. Because a significant share of patients at many hospitals are referred from hospitals outside the county, our indicators are constructed at the level of metropolitan statistical areas (MSA). At the MSA level, leakage still occurs (particularly patients from outlying rural areas receiving care hospitals in the MSA for intensive procedures), but it is relatively modest. The vast majority of patients treated in an MSA come from that MSA; and the vast majority of residents in the MSA receive treatment there. With more detailed information on patient residence, richer and more accurate area indicators could be constructed using the definitions applied in this report. Areas outside of any MSA were examined separately and on a county level.

Although these quality indicators are area-based, an important role remains for hospital-level measures of procedures or ambulatory care-sensitive admissions. If an area is found to

have unusually high procedure rates, then the hospitals that contribute substantially to those rates, and more specifically the population served by those hospitals, represent a natural focus for efforts to understand why rates are high and possibly to reduce them. Similarly, if an area is found to have unusually high rates of potentially avoidable admissions, then the patient populations treated by hospitals with a relatively large share of these admissions might be a good focus for interventions to understand and reduce hospitalization rates.

Organized by evidence level, Tables 11-12 summarize our literature review and empirical analysis for each of the recommended area indicators. All of these indicators have been evaluated in at least a limited number of previous studies, and all performed relatively well in our empirical evaluations.

**Utilization Indicators (area-level only)**

*Precision.* All of the recommended area-level utilization indicators vary substantially across MSAs, by rates that seem far larger than can be explained by plausible differences in area health characteristics. For each of these indicators, there is some problem with distinguishing true rate differences from random variations. However, the noise problems are much smaller than exist with the hospital level measures, because the numerators and denominators are generally much larger. Thus, for the most part, it is possible to obtain reliable estimates of area rates without the use of relatively complex statistical methods to account for random noise.

*Minimum bias.* Many factors other than differences in hospital practices may influence procedure use, and vary systematically across areas. Whereas some of these factors seem appropriate to exclude from risk adjustment (i.e. differences in rates resulting from differences in insurance coverage), others seem inappropriate (i.e. differences in clinical risk factors and health behaviors). In general, age and sex adjustment has little impact on the area measures, and because existing measures are incomplete, it is not feasible to adjust for differences in health status. (It is possible to adjust area rates using area characteristics, e.g. poverty rate. However, such analyses require careful attention to methodological issues such as the so-called "ecological fallacy.") Many of these influences on health care use are also associated in a complex way with differences in socioeconomic status. We note where previous studies permit some conclusions about whether these factors have relatively modest or substantial effects on area rates. In general, while a range of environmental, socioeconomic, and other factors have been shown to influence area rates, a substantial part of the variation in rates across areas is unexplained by all of these factors. One additional source of bias for the PTCA rate measure is the performance of some procedures on an outpatient basis. However, the share of such procedures is relatively small (less than 10%).

*Construct validity.* For most of the area indicators, previous studies have documented moderate or high utilization of the procedure for indications judged by experts to be inappropriate or of questionable value. One exception is CABG rates, which vary substantially across US areas and are much higher than in many other countries, but (at least in the areas studied so far) do not seem to be associated with substantial rates of inappropriate use (though rates of use for indications of uncertain value are high) and which may be underused in some patient subgroups. Most studies have not linked high procedure rates to higher rates of inappropriate use. For this reason, the area utilization rates are proxy measures for inappropriate

utilization only. Since the rate of inappropriate use appears not to vary with procedure rates, areas with higher rates have a higher raw volume of inappropriate procedures, than areas with lower rates.

*Fosters true quality improvement.* The studies on appropriateness of these procedures suggest that lowering procedure rates would primarily reduce inappropriate and low-value uses, leading to important benefits for the efficiency of medical care and possibly for patient outcomes. However, little direct evidence exists to date on this hypothesis. To achieve procedure reductions, it also seems likely that the high area rate must be parsed into contributions by specific hospitals. This is straightforward with HCUP data: the numerator of the area rate measures can be divided up into shares attributable to specific hospitals.

*Further investigation.* For most of the area utilization indicators, detailed clinical guidelines exist for judging the appropriateness of procedure use in specific cases. Such guidelines can be applied to sample cases from hospitals that make large contributions to high area rates, and to help identify specific opportunities for lowering rates, thereby providing more convincing evidence that no clinical harm would result. Information on patient residence could be used to identify and exclude patients from outside the area, and could also be used to provide a "proxy" (based on zip code) of patient income and other characteristics of the hospital area that may influence rates.

**ACSC Indicators**

To a large extent, the same issues described for the area utilization indicators also apply to the ACSC indicators.

*Precision.* All of the ACSC indicators are measured relatively precisely, and all involve serious complications that are at least somewhat common. However, methods to eliminate the effects of random noise on estimated rates are likely to be helpful, especially for the measures that are somewhat less common than the procedure utilization measures described above.

*Minimum bias.* All of the factors that may influence area utilization rates also influence area ACSC rates. In addition, some of the rates are substantially influenced by environmental conditions (e.g., COPD and pediatric asthma). For some indicators, differences in socioeconomic status have been shown to explain a substantial part – perhaps most – of the variation in ACSC rates across areas. However, this relationship is often used in the literature as proof of the validity of these conditions. The complexity of the relationship between SES and ACSC rates, make it difficult to delineate how much of the observed relationships are due to true access to care difficulties in potentially underserved populations, or from other patient characteristics, unrelated to quality of care, that vary systematically by SES. Finally, for some of the indicators, patient preferences and hospital capabilities for inpatient or outpatient care might explain variations in hospitalization rates.

*Construct validity.* In general, studies have shown that better outpatient care (including, in some cases, adherence to specific evidence-based treatment guidelines) can reduce patient complication rates, including the complications leading to ACSC admissions. For ruptured

appendix, hospital care in emergency departments (specifically, time to treatment) also appears to influence the rate of rupture; thus, this ACSC indicator has a component related to hospital rather than ambulatory care. Empirically, most of the ACSC rates are correlated with each other, suggesting that common underlying factors influence many of the rates.

*Fosters true quality improvement.* Despite the relationships demonstrated at the patient level between higher-quality ambulatory care and lower rates of admission with subsequent complication, there is generally little evidence on whether improvements in access to high-quality care can reduce ACSC hospitalization rates in an area. Such relationships are difficult to elucidate, because of the many intervening factors that also affect ACSC rates as noted above. On the other hand, there is also little evidence that use of these quality indicators would have any undesirable effects on hospital activities. Using HCUP data to identify the hospital patient populations making the largest contributions to area rates might also provide some insights into causes and potential responses to high ACSC rates, as those populations within those hospital service areas could be targeted for study or for intervention.

*Further investigation.* 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., pneumonia, dehydration). For these conditions, complementary analyses of data from the Medicare program, which include longitudinal records and information on outpatient care, can provide further insights regarding high area rates and whether or not they are associated with less use of outpatient care. Even though HCUP data are less detailed in some respects, 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 importance of "leakage" in and out of MSAs on apparent rate differences, and on the effects of socioeconomic and other area characteristics on rates. In addition, information on outpatient care for ACSCs can provide evidence on whether some of the admissions might have been avoidable (e.g., ruptured appendix, if patients with appendicitis are treated and released and then return with persistent symptoms) and on whether hospitals and areas differ in their ability to manage some of the ACSCs on an outpatient basis.

### Summary of Empirical Evaluation of Precision and Bias for Area Indicators

Tables 20 and 22 summarize the empirical evaluations of precision (including variation across areas) and bias (with respect to age and sex differences) for the recommended area utilization rate and ACSC indicators. Empirical findings on the performance of specific indicators are presented in detail in Appendix 7 and are summarized in Tables 11-12.

## Grouping of Indicators

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 the

performance of a quality indicator alone, including random variation. However, consistent good or bad performance on several related indicators is more convincing of a true underlying difference in performance, as it is more unlikely that such a pattern could arise from random or unsystematic events. Looking at groups of indicators together, therefore, is likely to provide a more complete picture of quality. While the HCUP indicators were not developed as "scales" of quality, meaning that one could calculate an overall quality score by plugging the indicators in an algorithm, they do group together in both by aspects of care and clinical domains. Generally, users would be most interested in either area performance or provider performance. However, in some cases, such as comprehensive medical groups, or policy regarding both hospital and area level quality, use of area and provider indicators together may be helpful. Possible indicator groupings are discussed below. Both provider and area level measures may be included, but users should keep in mind that these two types of indicators were designed to measure slightly different aspects of care.

Indicators are already grouped according to indicator type. This grouping can be used for more than simple organization. Indicators in similar groupings are often to designed to measure the same aspect of care, whether it be the outpatient access to care for ACSC conditions, or the healthcare utilization rates of utilization indicators. Thus examining indicators of like type may be useful in gaining a more complete picture of quality.

In the case of the ACSC indicators, many of the indicators were actually developed as part of a set, designed to comprehensively examine access to care. These indicators have been most often validated as a set, the evidence for which is set forth in the detailed write-up on ACSC indicators found in the last section of the results chapter of this report. In this case, use together as a set may be particularly ideal, since the evidence for some of these indicators alone is unclear.

In contrast, many of the mortality indicators have not been developed together, although they have been studied together. Using factor analysis, we found that medical mortality indicators in particular tend to be related to each other, meaning that providers with high rates for one condition also tended to have high rates for another condition. The pattern for post-procedural mortality is less clear, though some procedures tend to be positively related to each other, though to a lesser extent than medical mortality measures. Since different surgeons, and in some cases different surgical teams perform different operations, these indicators may vary more independently. Nevertheless, examining post-surgical indicators together may aid in identifying problems in overall surgical quality that is not procedure dependent, if such a problem existed.

The remaining utilization indicators all examine inappropriate use of procedures. Using factor analysis we found that area utilization indicators were related to each other, meaning areas with high utilization rates for one procedure tended to have high rates for the other three procedures. The provider level indicators however tended to follow another pattern. Utilization of technical procedures of bilateral catheterization and laparoscopic cholecystectomy tended to be negatively correlated, as the quality relationship would suggest, but not as strongly related to more established procedures like cesarean section, VBAC or incidental appendectomy. The latter three indicators were predictably related to each other.

Further grouping of indicators can be based on clinical domain. Though we did not perform formal analyses of the construct validity of these groupings, they have adequate face validity in that similar physicians and health care teams tend to provide care for the conditions in each grouping. For instance cardiologist or cardiovascular surgeons tend to provide care for all

of the cardiovascular indicators. Example groupings are suggested below, although further research on the validity of these grouping is needed.

Section 3.A. Summary of Results

**Table 8. Example clinical groupings**

| Domain name | Indicator | |
|---|---|---|
| *Cardiovascular care* | Provider-level | CABG volume |
| | | Pediatric heart surgery volume |
| | | PTCA volume |
| | | AMI mortality |
| | | Congestive heart failure mortality |
| | | CABG mortality |
| | | Pediatric heart surgery mortality |
| | Area-level | CABG area rate |
| | | PTCA area rate |
| | | Angina admission rate |
| | | CHF admission rate |
| *Cerebrovascular care* | Provider level | Carotid endarterectomy rate |
| | | Stroke mortality |
| *Other vascular care* | Provider level | AAA repair volume |
| | | AAA repair mortality |
| *Geriatric care* | Provider level | Incidental appendectomy in the elderly rate |
| | | Hip fracture mortality |
| | | Hip replacement mortality |
| | | Age stratified rate for pneumonia mo rtality |
| | | Age stratified rate for stroke mortality |
| | | Age stratified rate for dehydration admission rate |
| | | Age stratified rate for pneumonia admission rate |
| | | Age stratified rate for hypertension admission rate |
| *Obstetric care* | Provider level | Cesarean section rate |
| | | VBAC rate |
| | | Low birth weight rate |
| *Pediatric care* | Provider level | Pediatric heart surgery volume |
| | | Pediatric heart surgery mortality rate |
| | Area level | Pediatric asthma admission rate |
| | | Pediatric gastroenteritis admission rate |
| *Care for chronic conditions* | Provider level | CHF mortality rate |
| | Area level | Asthma admission rate (adult and pediatric) |
| | | COPD admission rate |
| | | CHF admission rate |
| | | Diabetes indicators |
| *Diabetes care* | Area level | Diabetes short term complications |
| | | Uncontrolled diabetes |
| | | Diabetes long term complications |
| | | Lower extremity amputation in diabetics |

## 3.A.3. Summary Tables of Evidence by Indicator

These tables are organized by indicator type, and either provider or area-level designation. All tables summarize the empirical and literature based evidence regarding the indicator. In addition, we make recommendations for using the indicator. One set of recommendations (HCUP users) are those that may be followed using the HCUP II software and administrative data like the HCUP database (e.g. use with other HCUP II QIs). Another set of recommendations (Future Investigations) suggests additional examinations, such as chart review or additional data collection, which may aid in the interpretation of the indicator.

**1) Table 9. Summary evidence table for provider-level volume indicators.** This table includes the following indicators (indicators #1 - #7): AAA repair volume, carotid endarterectomy volume, CABG volume, esophageal resection volume, pancreatic resection volume, pediatric heart surgery volume, PTCA volume.
**2) Table 10. Summary evidence table for provider-level utilization indicators.** This table includes the following indicators (indicators #8 - #12): Cesarean section delivery rate, incidental appendectomy among elderly rate, bilateral catheterization rate, VBAC rate, laparoscopic cholecystectomy rate.
**3) Table 11. Summary evidence table for area-level utilization indicators.** This table includes the following indicators (indicators #13 - #16): CABG rate, Hysterectomy rate, laminectomy rate, PTCA rate.
**4) Table 12. Summary evidence table for area-level ACSC indicators.** This table includes the following indicators (indicators #17 - #31): Dehydration admission rate, bacterial pneumonia admission rate, urinary infection admission rate, perforated appendix rate, angina admission rate, adult asthma admission rate, COPD admission rate, CHF admission rate, diabetes short-term complication admission rate, uncontrolled diabetes admission rate, diabetes long-term complication admission rate, hypertension admission rate, low birth-weight rate, pediatric asthma admission rate, and pediatric gastroenteritis admission rate.
**5) Table 13. Summary evidence table for provider-level mortality indicators.** This table includes the following mortality indicators (indicators #32 - #44): AMI, CHF, GI hemorrhage, hip fracture, pneumonia, acute stroke, AAA repair, CABG, craniotomy, esophageal resection, hip replacement, pancreatic resection and pediatric heart surgery.

**Table 9. Summary evidence table for volume indicators**

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| **Volume** | *Abdominal aortic aneurysm repair volume (#1)*<br><br>Higher volume may represent better quality care. | • *Precision:* Because unruptured AAA repair is considerably less common than many procedures for which volume-outcome relationship has been established (e.g., PTCA), the number of procedures performed in one year may not provide an accurate guide to hospital experience.<br>• *Minimum bias:* No bias due to outpatient procedures.<br>• *Construct validity:* Relationship between volume and outcome is well established, although the only study that used clinical data for risk-adjustment found no volume effect. No clear evidence on appropriate volume thresholds. Volume-outcome relationship for hospital volume may be weaker than for provider (operator) volume, among ruptured aneurysms.<br>• *Fosters quality improvement:* Large percentage of procedures performed in low-volume hospitals, suggesting that shifting care to high-volume centers could lead to a relatively large reduction in mortality from AAA repair.<br>• *Prior use*: This indicator is suggested in the literature, and the Pacific Business Group on Health notes that in lieu of other data, volumes for AAA may substitute as a measure of quality. | • *Precision:* Somewhat less common procedure; raw volume is moderately precise, with some consistency from year to year.<br>• *Minimum bias:* Risk adjustment not applicable.<br>• *Construct validity:* Correlated with AAA repair mortality, mortality for other procedures.<br>• *Fosters quality improvement:* 85% of AAA repair patients underwent aneurysmectomy at hospitals with volume greater than threshold 1 (10/yr); 43% underwent the procedure at hospital with volume greater than threshold 2 (32/yr). (Thresholds are based on cutoffs suggested in the literature). | *HCUP Data:*<br>• Proxy measure – "noisy" correlate of quality; not direct measure.<br>• Consider volume from more than one year if possible, as providers that appear to have volumes below desirable threshold in one year may be above threshold if volume is averaged over several years.<br>• Consider in conjunction with other procedure mortality indicators.<br><br>*Further Investigation:*<br>• Analysis of appropriateness of procedures is useful complement to hospital volume measure, although perhaps less important than for procedures with more elective indications.<br>• Volume-outcome relationships justifying any volume guidelines should be reevaluated periodically.<br>• Operator rates within hospitals should be examined if possible. |
| **Volume** | *Carotid endarterectomy volume (#2)*<br><br>Higher volume | • *Precision:* Endarterectomy volume is precisely measured with administrative data.<br>• *Minimum bias:* No bias due to outpatient procedures. | • *Precision:* Common procedure; raw volume is relatively precise, with much consistency from year to year.<br>• *Minimum bias:* Risk adjustment not applicable. | *HCUP Data:*<br>• Proxy measure – "noisy" correlate of quality; not direct measure.<br>• Consider in conjunction with endarterectomy mortality, other |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | may represent better quality care. | outpatient procedures.<br>• *Construct validity:* Relationship between volume and outcome is well established and apparently not due to patient case-mix differences. Volume-outcome relationship for hospital volume may be weaker than for surgeon provider (operator) volume, although different studies have produced conflicting results.<br>• *Fosters quality improvement:* Because a large percentage of procedures are performed by low-volume centers, many deaths and postoperative strokes may be preventable by shifting cases to high-volume centers. But imposing volume requirements might encourage more performance of procedures on asymptomatic patients, for whom the risk is low but the benefit is marginal.<br>• *Prior use*: This indicator is suggested in the literature. | applicable.<br>• *Construct validity:* Correlated with volume and mortality for other procedures, such as CABG.<br>• *Fosters quality improvement:* 78% of endarterectomy patients underwent endarterectomy at hospitals with volume greater than threshold 1 (50/yr); 51% underwent endarterectomy at hospital with volume greater than threshold 2 (101/yr). (Thresholds are based on cutoffs suggested in the literature). | endarterectomy mortality, other procedure mortality indicators.<br><br>*Further Investigation:*<br>• Analysis of appropriateness of procedures is useful complement to hospital volume measure, because of incentives for hospitals to increase volume.<br>• Volume-outcome relationships justifying any volume guidelines should be reevaluated periodically.<br>• Operator rates within hospitals should be examined if possible. |
| **Volume** | *Coronary artery bypass graft (CABG) volume (#3)*<br><br>Higher volume may represent better quality care. | • *Precision:* CABG is a common procedure, precisely measured with administrative data.<br>• *Minimum bias:* No bias due to outpatient procedures.<br>• *Construct validity:* Relationship between volume and outcome has been documented in many but not all studies, although some studies did not have detailed risk adjustment. The most recent studies, and studies using clinical data for risk-adjustment, suggest a non-significant association between volume and outcome. Volume-outcome relationship for hospital volume may be weaker than | • *Precision:* Common procedure; raw volume is precise, with high consistency from year to year. Additional statistical methods not needed.<br>• *Minimum bias:* Risk adjustment not applicable.<br>• *Construct validity:* Correlated with CABG mortality, mortality for other procedures.<br>• *Fosters quality improvement:* 98% of CABG patients underwent CABG at hospitals with volume greater than threshold 1 (100/yr); 91% underwent CABG at hospital with volume greater than threshold 2 (200/yr). | *HCUP Data:*<br>• Proxy measure – "noisy" correlate of quality; not direct measure.<br>• Consider in conjunction with CABG mortality, other procedure mortality indicators, and other indicators of cardiac care quality.<br><br>*Further Investigation:*<br>• Analysis of appropriateness of procedures is useful complement to hospital volume measure, because of incentives for hospitals to increase volume.<br>• Volume-outcome relationships justifying any volume guidelines |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | for surgeon volume, although the surgeon effect has also diminished over time. Higher-volume hospitals tend to perform more procedures on low-risk patients, who may benefit less from revascularization, though there is no evidence that this has occurred.<br>• *Fosters quality improvement:* Relatively few patients undergo CABG at low volume centers in some states; the percentage of avoidable deaths may vary substantially across states. But imposing volume requirements might encourage more performance of procedures on low-risk patients.<br>• *Prior use*: This indicator is suggested in the literature, and recommendations of minimum volumes to maintain proficiency have been recommended by professional organizations. | (Thresholds are based on cutoffs suggested in the literature). | should be reevaluated periodically.<br>• Operator rates within hospitals should be examined if possible. |
| **Volume** | *Esophageal resection volume (#4)*<br><br>Higher volume may represent better quality. | • *Precision:* Because esophageal resection is relatively uncommon, the number of procedures performed in one year may be a noisy indicator of hospital experience.<br>• *Minimum bias:* No bias due to outpatient procedures.<br>• *Construct validity:* Although a strong volume-outcome relationship has been documented, it is based on only three studies. No clear evidence on appropriate volume thresholds.<br>• *Fosters quality improvement:* Large percentage of procedures performed in low-volume hospitals where mortality appears much higher; however, | • *Precision:* Uncommon procedure; raw volume is less precise, with some variation from year to year.<br>• *Minimum bias:* Risk adjustment not applicable.<br>• *Construct validity:* Correlated with esophageal resection mortality and pancreatic section mortality.<br>• *Fosters quality improvement:* 40% of esophageal resection patients underwent resection at hospitals with volume greater than threshold 1 (when averaged over two years) (6/yr); 74% underwent the procedure at hospital with volume greater than threshold 2 (7/yr). (Thresholds are | *HCUP Data:*<br>• Proxy measure – "noisy" correlate of quality; not direct measure.<br>• Consider volume from more than one year if at all possible, as providers that appear to have volumes below desirable threshold in one year may be above threshold if volume is averaged over several years.<br>• Consider in conjunction with pancreatic resection mortality.<br><br>*Further Investigation:*<br>• Volume-outcome relationships justifying any volume guidelines should be reevaluated periodically. |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | relatively little gain in patient mortality is possible by shifting care to high-volume centers due to small number of surgeries performed. Some patients with poor prognosis may have to travel large distances for treatment by a high-volume center.<br>• *Prior use*: This indicator is suggested in the literature. | based on cutoffs suggested in the literature). | • Operator rates within hospitals should be examined if possible. |
| **Volume** | *Pancreatic resection volume (#5)*<br><br>Higher volume may represent better quality. | • *Precision:* Because pancreatic resection is relatively uncommon, the number of procedures performed in one year may be a noisy indicator of hospital activity.<br>• *Minimum bias:* No bias due to outpatient procedures.<br>• *Construct validity:* Relationship between volume and outcome is well established and apparently not due to patient case-mix differences. No clear evidence on appropriate thresholds.<br>• *Fosters quality improvement:* Large percentage of procedures performed in low-volume hospitals where mortality appears much higher; however, relatively little gain in patient mortality is possible by shifting care to high-volume centers due to small number of surgeries performed. Some patients with poor prognosis may have to travel large distances for treatment by a high-volume center.<br>• *Prior use*: This indicator is suggested in the literature. | • *Precision:* Uncommon procedure; raw volume is relatively imprecise, with considerable variation from year to year.<br>• *Minimum bias:* Risk adjustment not applicable.<br>• *Construct validity:* Correlated with pancreatic resection mortality and esophageal resection mortality.<br>• *Fosters quality improvement:* 30% of pancreatic resection patients underwent resection at hospitals with volume greater than threshold 1 (when averaged over two years) (10/yr); 27% underwent the procedure at hospital with volume greater than threshold 2 (11/yr). (Thresholds are based on cutoffs suggested in the literature). | *HCUP Data:*<br>• Proxy measure – "noisy" correlate of quality; not direct measure.<br>• Consider volume from more than one year if at all possible, as providers that appear to have volumes below desirable threshold in one year may be above threshold if volume is averaged over several years.<br>• Consider in conjunction with esophageal resection mortality.<br><br>*Further Investigation:*<br>• Volume-outcome relationships justifying any volume guidelines should be reevaluated periodically.<br>• Operator rates within hospitals should be examined if possible. |
| **Volume** | *Pediatric heart surgery volume (#6)* | • *Precision:* Heart surgery volume is precisely measured with administrative data; but hospital procedure volume is highly skewed. | • *Precision:* Because of skewed distribution of procedure volume, high consistency from year to year in whether a hospital is above desirable | *HCUP Data:*<br>• Proxy measure – "noisy" correlate of quality; not direct measure. |

Section 3.A.3. Summary Table of Results: Volume Indicators

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | Higher volume may represent better quality care. | • *Minimum bias:* No bias due to outpatient procedures.<br>• *Construct validity:* Relationship between volume and outcome is well established, although based on only three studies.<br>• *Fosters quality improvement:* Skewed distribution of volume makes it difficult for hospital to increase volume in order to exceed volume threshold. Because pediatric heart surgery is already largely regionalized, relatively little outcome improvement expected from increased use of high-volume centers.<br>• *Prior use*: This indicator is suggested in the literature. | volume thresholds.<br>• *Minimum bias:* Risk adjustment not applicable.<br>• *Construct validity:* Correlated with pediatric heart surgery mortality.<br>• *Fosters quality improvement:* 74% of pediatric heart surgery patients underwent surgery at hospitals with volume greater than the threshold (100/yr). (Thresholds are based on cutoffs suggested in the literature). | • Consider in conjunction with other procedure mortality indicators.<br><br>*Further Investigation:*<br>• Analysis of appropriateness and timing of procedures is useful complement to hospital volume measure.<br>• Volume-outcome relationships justifying any volume guidelines should be reevaluated periodically.<br>• Operator rates within hospitals should be examined if possible.<br>• Regionalization of procedures is already well established in most states and is thus clearly feasible. |
| **Volume** | *PTCA volume (#7)*<br><br>Higher volume may represent better quality care. | • *Precision:* PTCA is a common procedure, precisely measured with administrative data.<br>• *Minimum bias:* A minority (7.6%) of PTCAs are performed on ambulatory basis, so it is possible that a hospital's PTCA volume will be modestly understated with observational data.<br>• *Construct validity:* Relationship between volume and outcome is well established, although the magnitude of the effect is small and adverse outcome rates are generally low. PTCA volume explains only a small share of variation in PTCA outcomes across hospitals.<br>• *Fosters quality improvement:* Deaths and post-procedural CABGs may be avoidable by shifting care to high volume centers. But imposing volume requirements might encourage more | • *Precision:* Common procedure; raw volume is precise, with high consistency from year to year. Additional statistical methods not needed.<br>• *Minimum bias:* Risk adjustment not applicable.<br>• *Construct validity:* Correlated with volume, mortality for other procedures, such as CABG.<br>• *Fosters quality improvement:* 96% of patients underwent PTCA at hospitals with volume greater than threshold 1 (200/yr); 77% underwent PTCA at hospital with volume greater than threshold 2 (400/yr). (Thresholds are based on cutoffs suggested in the literature). | *HCUP Data:*<br>• Proxy measure – "noisy" correlate of quality, not a direct measure.<br>• Consider in conjunction with PTCA mortality, other procedure mortality indicators, and other indicators of cardiac care quality.<br><br>*Further Investigation:*<br>• Consider volume of outpatient PTCA procedures (relatively rare).<br>• Analysis of appropriateness of procedures is useful complement to hospital volume measure, because of incentives for hospitals to increase volume.<br>• Volume-outcome relationships justifying any volume guidelines should be reevaluated periodically.<br>• Operator rates within hospitals should be examined if possible. |

Section 3.A.3. Summary Table of Results: Volume Indicators

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | performance of procedures on low-risk patients.<br>• *Prior use*: This indicator is suggested in the literature, and recommendations of minimum volumes to maintain proficiency have been recommended by professional organizations. | | be examined if possible. |

**Table 10. Summary evidence table for provider level utilization indicators**

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| **Utilization** | *Cesarean section delivery rate (#8)*<br><br>A lower c-section rate may represent better quality. | • *Face validity:* C-section is a common procedure, rates vary substantially across providers, and medical advisory groups have made significant rate reductions an important policy goal (current rate is around 20%; Healthy People 2010 target is 15%).<br>• *Precision:* Common procedure, large denominator population at most hospitals, and considerable variation across providers indicate that true differences in provider rates can be distinguished.<br>• *Minimum bias:* Though some studies have suggested that risk adjustment does not substantially alter hospital rankings, other studies have found larger effects (though none controlled for random variations in comorbidity).<br>• *Construct validity:* Limited evidence on whether lower C-section rates lead to better patient outcomes, though costs are lower.<br>• *Fosters quality improvement:* Lower C-section rates or higher VBAC rates may lead to worse outcomes if procedures are deterred in appropriate candidates. No evidence for this phenomenon in states that have used C-Section rates as a quality indicator.<br>• *Prior use*: Cesarean section delivery rate is one of the most commonly used indicators. It is currently used by University Hospital Consortium, Florida Agency for Health Care Administration, Greater New York | • *Precision:* Indicator demonstrates substantial variation in performance across providers. Most of the observed variation reflects true differences in cesarean section rates.<br>• *Minimum bias:* Risk adjustment based on mother age only is feasible, because maternal-infant data cannot be reliably linked in HCUP and because APR-DRG severity does not vary for C-section. In the case of age adjustment, there was no significant impact.<br>• *Construct validity:* Performance on indicator is inversely related to VBAC rate – VBAC use is important component of variations in C-section rates – and positively related to incidental appendectomy rate. | *HCUP Data:*<br>• High C-section rates are likely associated with low VBAC rates, so that VBAC may be a useful target for interventions to reduce high rates of use of C-sections.<br>*Further Investigation:*<br>• Providers should compare rates with regional and national rates, to determine whether hospital rates differ from area and national norms.<br>• Indications for appropriate use of C-section are difficult to observe with discharge data; consider use of birth records or chart data to investigate whether high or low rates reflect a true quality problem.<br>• If maternal and child discharge records can be linked, useful risk adjusters can be developed from the infant discharge data.<br>• Because appropriate rate has not been definitively established, comparison to regional and national norms rather than a "gold standard" rate may be necessary. |

Section 3.A.3. Summary Table of Results: Utilization Indicators

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | Hospital Association, Maryland Quality Indicator Project, Michigan Hospital Association, Pacific Business Group on Health, United Health Care, Cleveland Health Quality Choice, IMSystem, Virginia Health Information, Washington State Community Health Information Partnership, and HealthGrades.com. It has been proposed for use by Colorado Health and Hospital. Cesarean section delivery rate is currently an HCUP I indicator. Repeat cesarean section delivery rate is used as an indicator by University Hospital Consortium. Primary cesarean section delivery rate is used as an indicator by National Center for Healthcare Statistics. | | |
| **Utilization** | *Incidental appendectomy in the elderly (#9)*<br><br>A lower incidental appendectomy rate may represent better quality. | • *Face validity:* Relatively strong consensus that this procedure is inappropriate, particularly for the elderly, due to low rate of appendicitis and increased potential for infection after abdominal surgery. Procedure is not widely used.<br>• *Precision:* Procedure is relatively infrequently used, (e.g., in less than 20% of routine cholecystectomies in Medicare beneficiaries in 1992), suggesting that QI may not be precisely measured in many hospitals.<br>• *Minimum bias:* Asymptomatic disease of appendix is rare, with no known risk factors. Therefore, unlikely that appropriate rate of incidental appendectomy would vary substantially across hospitals.<br>• *Construct validity:* No clear evidence | • *Precision:* Indicator demonstrates adequate variation in performance across providers. Some of the observed variation reflects true differences in incidental appendectomy rate, though some reflects unsystematic variation. Multivariate smoothing is particularly beneficial to this indicator.<br>• *Minimum bias:* Provider rankings and absolute performance are not affected by APR-DRG risk adjustment.<br>• *Construct validity:* Positively associated with C-section rate, and negatively associated with VBAC rate. Meaning providers with high cesarean section rates and/or low VBAC rates, also tend to have high incidental appendectomy rates. | *HCUP Data Use:*<br>• Though incidental appendectomy varies adequately across providers, some of the variation is due to unsystematic variation; "smoothed" estimates help avoid this problem.<br>• Consider in conjunction with other surgical utilization measures, particularly those related to obstetrics and abdominal surgery.<br><br>*Further Investigation:*<br>• Little evidence that supplemental data (e.g., chart reviews) would substantially alter conclusions about inappropriate use of incidental appendectomy. |

Section 3.A.3. Summary Table of Results: Utilization Indicators

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | that hospitals with generally better surgical practices are less likely to perform incidental appendectomy.<br>• *Fosters quality improvement:* It is possible that use of this QI could cause hospitals to under-report cases of incidental appendectomy, rather than truly reducing their incidental appendectomy rate. Removal of a truly inflamed appendix is clearly indicated, and so is unlikely to be deterred.<br>• *Prior use*: Currently an HCUP I indicator. | | |
| **Utilization** | *Bi-lateral cardiac catheterization (cath) rate (#10)*<br><br>Lower bi-lateral cath rate may represent better quality. | • *Precision:* Catheterization is a common procedure and bilateral cath is often performed, suggesting that precise estimates can be obtained for most hospitals.<br>• *Minimum bias:* Conditions for which right-heart cath may be appropriate are not fully captured in administrative data. But incidence of these conditions among cath patients is very low, suggesting that they are likely to lead to little bias at the provider level. A fair number of catheterizations are performed on an outpatient basis, leading to some bias for this indicator.<br>• *Construct validity:* Indications for bilateral cath are relatively rare. Procedure has significant incremental cost, very low yield, and some associated morbidity in many patients receiving it.<br>• *Fosters quality improvement:* Seems unlikely to be subject to much gaming, as patients who are appropriate | • *Precision:* Indicator demonstrates very substantial variation in performance across providers. Almost all of the observed variation reflects true differences in bilateral catheterization rates.<br>• *Minimum bias:* Provider rankings and absolute performance are not affected by APR-DRG risk adjustment.<br>• *Construct validity:* High bilateral cath rate associated with better performance on CABG mortality, and on laparoscopic cholecystectomy rate. | *Further Investigation:*<br>• As a fair number of heart catheterizations are performed on an outpatient basis, examination of outpatient data in addition to inpatient data may provide a more complete picture of performance on this indicator.<br>• Check clinical charts for appropriateness of procedures on patients undergoing procedure at high rate providers. |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | candidates are only a small share of all cath patients. Financial incentives may encourage performance.<br>• *Prior use*: Bi-lateral catheterization has been targeted by HCFA as a potentially misused procedure. | | |
| **Utilization** | *VBAC rate (#11)*<br><br>Higher VBAC rate may represent better quality. | • *Precision:* Relatively common procedure, suggesting that accurate estimates can be obtained for hospitals performing a significant number of deliveries.<br>• *Minimum bias:* Risk factors reducing the appropriateness of VBAC are not well captured in administrative data. However, these contraindications are relatively rare. Occurrence of prior C-section may be undercoded at some hospitals, biasing the measure upwards. Patient and provider preference may substantially impact rate.<br>• *Construct validity:* Approximately 1/3 of patients prefer repeat c-section to VBAC.<br>• *Fosters quality improvement*: Very high rate may not be clinically prudent, as some complications, such as uterine rupture may increase with VBAC.<br>• *Prior use*: VBAC is one of the most widely used indicators. Users include IMSystem, Maryland QI project, Michigan Hospital Association, HealthGrades.com, Cleveland Health Quality Choice Clinic, HCUP, University Hospital Consortium. VBAC rate was also selected as a JCAHO core measure. | • *Precision:* Indicator demonstrates very substantial variation in performance across providers. Most of the observed variation reflects true differences in VBAC rate.<br>• *Minimum bias:* Risk adjustment based on mother age only is feasible, because maternal-infant data cannot be reliably linked in HCUP and because APR-DRG severity does not vary for VBAC. In the case of age adjustment, there was no significant impact.<br>• *Construct validity:* High rate associated with better performance on cesarean section rate and incidental appendectomy rate. | *Further Investigation:*<br>• Due to possible underreporting of prior C-sections, low-rate providers should examine charts for systematic failure to document prior C-sections.<br>• Because appropriate rate has not been definitively established, comparison to regional and national norms rather than a "gold standard" rate may be necessary.<br>• Can be used in conjunction with measures of neonatal complications and length of stay, to provide further evidence on whether low or high rate reflects quality problems.<br>• If maternal and child discharge records can be linked, useful risk adjusters can be developed from the infant discharge data. |

Section 3.A.3. Summary Table of Results: Utilization Indicators

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| **Utilization** | *Laparoscopic cholecystectomy (LC) (#12)*<br><br>A higher percentage of LC (compared to all cholecystectomies) may represent higher quality. | • *Face validity:* Improved outcomes and lower costs associated with laparoscopic cholecystectomy have been well-demonstrated, so that this procedure is standard of care for uncomplicated cholecystectomy patients. Lower rate may reflect less adoption of effective new innovations, and/or poorer technical skill in performing new procedures.<br>• *Precision:* Relatively common procedure, so reasonably precise measures appear feasible.<br>• *Minimum bias:* Though many risk factors influencing the appropriate use of laparoscopic cholecystectomy are available in discharge abstracts, not all factors are captured (e.g. patient factors such as difficult anatomy are not well-captured). A substantial number of procedures are performed on an outpatient basis.<br>• *Construct validity:* Though considerable evidence indicating near-universal adoption of procedure, little evidence exists on the extent to which higher LC rates reflect higher quality of care.<br>• *Fosters quality improvement:* Providers could improve rankings by performing more procedures on low-risk, low-benefit patients (e.g., those with only mild GI symptoms).<br>• *Prior use*: Currently an HCUP I indicator. | • *Precision:* Indicator demonstrates very substantial variation in performance across providers. Most of the observed variation reflects true differences in laparoscopic cholecystectomy rate.<br>• *Minimum bias:* Overall very little impact of age-sex risk adjustment; additional risk adjustment not feasible with APR-DRGs.<br>• *Construct validity:* Higher LC rate is associated with significantly better performance bilateral catheterization rate and CABG mortality; may reflect more effective adoption of recent clinical practice developments.<br>• *Fosters quality improvement:* Better performance on measure is significantly associated with higher area rate of procedure. | *HCUP Data:*<br>• Contrary to common perception, substantial differences in the LC quality indicator between providers exist.<br><br>*Further Investigation:*<br>• Consider in conjunction with overall area cholecystectomy rate (laparoscopic plus open), as indicator of whether too many low-benefit procedures are being performed.<br>• Can be used in conjunction with measures of length of stay and postoperative complications, to provide additional evidence on surgeon quality in performing laparoscopic procedures.<br>• Though discharge diagnoses permit some adjustment for risk differences, further evaluation could involve chart review to determine whether hospitals differ in prevalence of low-risk, mild-symptom cases and/or cases with higher risk of LC complications. |

**Table 11. Summary evidence table for area-level utilization indicators.**

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| **Utilization, area level** | *Coronary Artery Bypass Graft (CABG) area rate (#13)*<br><br>Lower rates may represent better quality of care. | • *Face Validity:* CABG rates vary substantially across geographic areas, in relation to the availability of cardiac surgery facilities. This variation may be associated with use of the procedure for clinical indications in which benefits have not been demonstrated.<br>• *Minimum bias:* CABG rate may be related to incidence of CAD, which may be in turn related to the age, smoking, hyperlipidemia, and other characteristics of the population. Although little evidence exists on whether such characteristics explain a significant component of area variation in CABG rates, this hypothesis seems implausible.<br>• *Construct validity:* Most studies suggest relatively little inappropriate use of CABG. A larger percentage of CABG procedures are performed for indications where benefits are uncertain; whether outcomes would be affected by reducing the use of such procedures is not clear.<br>• *Fosters true quality improvement:* Little evidence on whether indicator use would selectively reduce inappropriate procedures or procedures of uncertain value.<br>• *Prior use:* Included in the previous HCUP I indicator set, and in the Dartmouth Atlas. | • *Precision:* Indicator demonstrates very substantial variation in performance across areas. Almost all of the observed variation reflects true differences in procedure rates.<br>• *Minimum bias:* Area rankings are substantially affected by age-sex risk adjustment, suggesting that patient characteristics influence performance on indicator. However, absolute performance is only moderately affected by risk adjustment.<br>• *Construct validity:* Areas with high CABG rates also tend to have higher rates of other procedures captured in area utilization indicators. | *HCUP Data:*<br>• Indirect measure of "quality," since area rates may not reflect inappropriate use.<br>• Due to the lack of evidence for the construct validity of this indicator, CABG mortality is recommended only for use with corresponding volume and mortality measures.<br>• Almost all of the apparent rate variation across areas reflects true differences; additional statistical methods are not needed to account for random noise.<br>• Risk adjustment with age-sex at minimum is recommended for this indicator, due to evidence of potential bias.<br>• If area rate is high, can use HCUP data to identify area hospitals making greatest contribution to area rate.<br><br>*Further Investigation:*<br>• Ideal rate of CABG has not been established, and probably varies with area demographic and health characteristics. Previous studies have developed standards for classifying indications for CABG as appropriate, questionable, and inappropriate. These standards can be applied to chart reviews in area hospitals to determine whether significant inappropriate or questionable use of CABG is |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | | | occurring.<br>• High area rate may be due in part to performance of procedures on patients referred from outside the area; hospital discharge data with information on patient residence can address whether such patients account for a significant share of area procedures. |
| **Utilization, area level** | *Hysterectomy area rate (#14)*<br><br>Lower rates may represent better quality care. | • *Face Validity:* Hysterectomy rates vary substantially across small geographic areas, with many procedures performed for mild to moderate symptoms of uterine bleeding, pelvic pain, or other conditions where the benefit of the procedure is unclear.<br>• *Precision:* Relatively common procedure, so that previous studies have identified significant systematic variation in hysterectomy rates across areas after accounting for random variation.<br>• *Minimum bias:* Hysterectomy rates differ by socioeconomic status and patient characteristics, but little evidence on the extent to which population differences can explain area rate variations.<br>• *Construct validity:* Wide variation across areas in the proportion of cases deemed inappropriate, although there is no evidence that inappropriate use explains the variation in utilization rates.<br>• *Fosters true quality improvement:* Little evidence on whether indicator use would selectively reduce | • *Precision:* Indicator demonstrates very substantial variation in performance across areas. Almost all of the observed variation reflects true differences in procedure rates.<br>• *Minimum bias:* Area rankings and absolute performance are not affected by age risk adjustment.<br>• *Construct validity:* Areas with high hysterectomy rates also tend to have higher rates of other procedures captured in area utilization indicators. | *HCUP Data:*<br>• Indirect measure of "quality," since area rates may not reflect inappropriate use.<br>• Almost all of the apparent rate differences across areas reflect true differences; additional statistical methods to account for random noise are relatively unimportant.<br>• Demographic adjustment of area rates has some effect on absolute area performance.<br>• If area rate is high, can use HCUP data to identify area hospitals making greatest contribution to area rate.<br><br>*Further Investigation:*<br>• Ideal rate of hysterectomy has not been established, and probably varies with area demographic and health characteristics. Previous studies have developed standards for classifying indications for hysterectomy as appropriate, questionable, and inappropriate. These standards can be applied to chart reviews in area hospitals to determine whether significant |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | inappropriate hysterectomies.<br>• *Prior use:* Included in the previous HCUP I indicator set, and in the Dartmouth Atlas. | | inappropriate use of hysterectomy is occurring.<br>• High area rate may be due to performance of procedures on patients referred from outside the area; hospital discharge data with information on patient residence can address whether such patients account for a significant share of area procedures. |
| **Utilization, area level** | *Laminectomy and/or spinal fusion area rate (#15)*<br><br>Lower rates may represent better quality of care. | • *Face Validity:* Laminectomy is a relatively common procedure for which very high geographic variation has been reported, particularly in studies on elderly patients using Medicare data. Many patients appear to undergo laminectomy for indications in which benefits have not been demonstrated involving symptoms for which benefits of procedure are unclear.<br>• *Minimum bias:* Differences in sociodemographic, economic, health, and other factors may influence laminectomy use. However, such factors appear to account for little of the large variation in procedure rates observed across areas.<br>• *Construct validity:* A substantial percentage of laminectomy procedures appear to be performed for inappropriate reasons, although there is no evidence that inappropriate use explains the variation in utilization rates.<br>• *Fosters true quality improvement:* Little evidence exists on whether indicator use would selectively reduce | • *Precision:* Indicator demonstrates substantial variation in performance across areas. Almost all of the observed variation reflects true differences in procedure rates.<br>• *Minimum bias:* Area rankings are somewhat affected by age-sex risk adjustment. However, risk adjustment does not change the absolute performance of areas substantially.<br>• *Construct validity:* Areas with high laminectomy rates tend to have high rates for area utilization indicators. | *HCUP Data:*<br>• Indirect measure of "quality," since area rates may not reflect inappropriate use.<br>• Almost all of the apparent rate variation across areas reflects true differences; additional statistical methods are not needed to account for random noise.<br>• Demographic adjustment of area rates has some effect on the relative ranking areas, particularly those with high rates.<br>• If area rate is high, can use HCUP data to identify area hospitals making greatest contribution to area rate.<br><br>*Further Investigation:*<br>• Ideal rate of laminectomy is unclear, and probably varies with area demographic and health characteristics. Previous studies have developed standards for classifying indications for laminectomy as appropriate, questionable, and inappropriate. These standards can be applied to |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | inappropriate procedures or procedures of uncertain value.<br>• *Prior use:* Included in the previous HCUP I indicator set, and in the Dartmouth Atlas. | | chart reviews in area hospitals to determine whether significant inappropriate or questionable use of laminectomy is occurring.<br>• High area rate may be due in part to performance of procedures on patients referred from outside the area; hospital discharge data with information on patient residence can address whether such patients account for a significant share of area procedures. |
| **Utilization, area level** | *PTCA area rate (#16)*<br><br>Lower rates may represent better quality care. | • *Face Validity:* Substantial variation in PTCA use across hospitals and areas, including use in many patients for whom clinical benefit has not been demonstrated in randomized trials.<br>• *Precision:* Common procedure, suggesting that area rate estimates should be relatively precise.<br>• *Minimum bias:* Demographic, socioeconomic, and health differences across area populations may contribute to differences in observed rates; however, observed variation appears far too large to be attributable primarily to such factors. Failure to capture outpatient procedures in HCUP data may also bias estimates; however, too few procedures are performed on an outpatient basis (<10%) to account for a substantial part of the variation across areas.<br>• *Construct validity:* Substantial and highly-variable percentage of PTCAs have been deemed inappropriate in different studies, suggesting that geographic variation in inappropriate | • *Precision:* Indicator demonstrates very substantial variation in performance across areas. Almost all of the observed variation reflects true differences in procedure rates.<br>• *Minimum bias:* Area rankings are substantially affected by age-sex risk adjustment, suggesting that patient characteristics influence performance on indicator. However, absolute performance is only moderately impacted.<br>• *Construct validity:* PTCA rate is positively related to other area utilization indicators. | *HCUP Data:*<br>• Indirect measure of "quality," since area rates may not reflect inappropriate use.<br>• Almost all of the apparent rate variation across areas reflects true differences; additional statistical methods are not needed to account for random noise.<br>• Risk adjustment with age-sex at minimum is recommended for this indicator, due to evidence of potential bias.<br>• If area rate is high, can use HCUP data to identify area hospitals making greatest contribution to area rate.<br><br>*Further Investigation:*<br>• Ideal rate of PTCA has not been established, is probably changing over time, and probably varies with area demographic and health characteristics. Previous studies have developed standards for classifying indications for PTCA as |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | and equivocal use may be substantial. No evidence that inappropriate use explains the variation in utilization rates.<br>• *Fosters true quality improvement:* Little evidence on whether indicator use would selectively reduce inappropriate PTCAs.<br>• *Prior use:* Included in the Dartmouth Atlas. | | appropriate, questionable, and inappropriate. These standards can be applied to chart reviews in area hospitals to determine whether significant inappropriate or questionable use of PTCA is occurring.<br>• High area rate may be due to performance of procedures on patients referred from outside the area; hospital discharge data with information on patient residence can address whether such patients account for a significant share of area procedures. |

**Table 12. Summary evidence table for area-level ACSC indicators.**

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| **Avoidable Hospi-talization ACSC** | *Dehydration (#17)*<br><br>Lower rates may represent better quality of care. | • *Face Validity:* Dehydration is a serious, potentially fatal condition in frail patients and patients with serious comorbid illnesses. Appropriate attention to fluid status can prevent the condition, and effective IV fluid therapy can avoid complications necessitating admission when it occurs.<br>• *Precision:* Somewhat common cause of hospital admission. Little evidence exists on the precision and variation associated with this indicator.<br>• *Minimum bias:* Little evidence exists on the importance of confounding differences in population risk factors. Differences in thresholds for admission of patients with dehydration (versus management on an ambulatory basis or through emergency room care) may also contribute to area rate differences.<br>• *Construct validity:* Little evidence exists on the construct validity of this indicator.<br>• *Fosters true quality improvement:* Use of this indicator might lead to higher thresholds of admission for patients with dehydration, but no evidence suggests that this would lead to worse patient outcomes.<br>• *Prior use:* Developed by Billings. | • *Precision:* Indicator demonstrates adequate variation in performance across areas. Most of the observed variation reflects true differences in admission rates.<br>• *Minimum bias:* Area rankings are not affected by age-sex risk adjustment. However, risk adjustment does change the absolute performance of areas somewhat.<br>• *Construct validity:* Areas with high rates of dehydration admissions also tend to have high rates of admissions for other ACSCs. | *HCUP Data:*<br>• Potentially an indirect measure of "quality," since area rates may not reflect quality of outpatient care.<br>• Considerable variations across areas, suggesting opportunities for quality improvement in care of patients at risk for dehydration.<br>• Accounting for measurable risk differences across areas somewhat affects relative area rankings.<br>• Consider in conjunction with other ACSC indictors of short-term complications of poor outpatient care, since all are generally positively correlated.<br>• Interventions to improve outpatient and emergency care can be targeted to patients treated by hospitals that admit relatively large shares of dehydration patients in an area.<br>*Further investigation:*<br>• Little previous literature to guide investigations of underlying causes of variations in dehydration admission rates.<br>• Examination of outpatient dehydration care may provide a more complete picture of care for this condition. This evaluation might include hospital emergency department care, which can lead to lower admission rates for patients with dehydration.<br>• Inpatient admissions plus emergency room visits for |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | | | dehydration may provide a more complete picture of the rate of clinically significant cases of dehydration in an area.<br>• Because most hospitalizations involve the elderly, Medicare data may be able to provide additional evidence on dehydration admission rates and their determinants, at least for areas with low Medicare managed care enrollment (and hence incomplete data). |
| **Avoidable Hospi-talization ACSC** | *Bacterial Pneumonia (#18)*<br><br>Lower rates may represent better quality of care. | • *Face Validity:* Vaccination for pneumococcal pneumonia in the elderly have been shown to be 45% effective in preventing admissions. Early diagnosis and effective management of bacterial respiratory infections on an ambulatory basis may also reduce admissions with pneumonia.<br>• *Precision:* Admissions for pneumonia are very common, suggesting that reasonably precise estimates of area rates should be feasible.<br>• *Minimum bias:* Little evidence exists on the importance of confounding differences in population risk factors. Differences in thresholds for admission of patients with pneumonia (versus management on an ambulatory basis or through emergency room care) may also contribute to area rate differences.<br>• *Construct validity:* Little evidence exists on the construct validity of this indicator.<br>• *Fosters true quality improvement:* | • *Precision:* Indicator demonstrates adequate variation in performance across areas. Almost all of the observed variation reflects true differences in admission rates.<br>• *Minimum bias:* Area rankings and absolute performance are somewhat affected by age-sex risk adjustment.<br>• *Construct validity:* Areas with high rates of bacterial pneumonia admissions also tend to have high rates of admissions for other ACSCs. | *HCUP Data:*<br>• Potentially an indirect measure of "quality," since area rates may not reflect quality of outpatient care.<br>• Considerable variations across areas, suggesting opportunities for quality improvement in care of pneumonia patients, including more prevention via higher vaccination rates.<br>• Accounting for measurable risk differences across areas somewhat affects relative area rankings.<br>• Consider in conjunction with other ACSC indictors of short-term complications of poor outpatient care, since all are generally positively correlated.<br>• Interventions to improve outpatient and emergency care for pneumonia can be targeted to patients treated by hospitals that admit relatively large shares of pneumonia patients in an area.<br>*Further investigation:*<br>• Little previous literature to guide |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | Use of this indicator might lead to higher thresholds of admission for pneumonia patients, but no evidence suggests that this would lead to worse patient outcomes.<br>• *Prior use:* Developed by Weissman. Related indicator of immunization preventable pneumonia developed by Billings and Weissman, and included in the previous HCUP I indicator set. | | investigations of underlying causes of variations in pneumonia admission rates.<br>• Examination of outpatient pneumonia care and area pneumococcal pneumonia vaccination rates may provide a more complete picture of care for this condition. This evaluation might include hospital emergency department care, which can lead to lower admission rates for patients with pneumonia.<br>• Inpatient admissions plus emergency room visits for pneumonia may provide a more complete picture of the rate of clinically significant cases of pneumonia in an area.<br>• Because most hospitalizations involve the elderly, Medicare data may be able to provide additional evidence on pneumonia admission rates and their determinants, at least for areas with low Medicare managed care enrollment (and hence incomplete data). |
| **Avoidable Hospi-talization ACSC** | *Urinary tract infection (UTI) (#19)*<br><br>Lower rates may represent better quality of care. | • *Face Validity:* Although uncomplicated urinary tract infections can be managed on an ambulatory basis, failure to treat with antibiotics or under- or inappropriate treatment can lead to more serious complications, including pyelonephritis and sepsis, which require hospitalization.<br>• *Precision:* Somewhat common cause of hospital admission. Little evidence | • *Precision:* Indicator demonstrates adequate variation in performance across areas. Most of the observed variation reflects true differences in admission rates.<br>• *Minimum bias:* Area rankings and absolute performance are somewhat affected by age-sex risk adjustment.<br>• *Construct validity:* Areas with high admission rates for urinary tract infections also tend to have high | *HCUP Data:*<br>• Potentially an indirect measure of "quality," since area rates may not reflect quality of outpatient care.<br>• Considerable variation across areas, suggesting opportunities for quality improvement in care of UTI patients.<br>• Accounting for measurable risk differences across areas somewhat affects relative area rankings. |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | exists on the precision and variation associated with this indicator.<br>• *Minimum bias:* Little evidence exists on the importance of confounding differences in population risk factors. Differences in thresholds for admission of patients with urinary tract infection (versus management on an ambulatory basis or through emergency room care) may also contribute to area rate differences.<br>• *Construct validity:* Little evidence exists on the construct validity of this indicator.<br>• *Fosters true quality improvement:* Use of this indicator might lead to higher thresholds of admission for patients with urinary tract infections, but no evidence suggests that this would lead to worse patient outcomes.<br>• *Prior use:* Developed independently by Billings and Weissman. Included in the UK National Health Service High Level Performance Indicators. | admission rates for other ACSCs. | • Consider in conjunction with other ACSC indictors of short-term complications of poor outpatient care, since all are generally positively correlated.<br>• Interventions to improve outpatient and emergency care can be targeted to patients treated by hospitals that admit relatively large shares of UTI patients in an area.<br>*Further investigation:*<br>• Little previous literature to guide investigations of underlying causes of variations in UTI admission rates.<br>• Examination of outpatient care of UTIs may provide a more complete picture of care for this condition. This evaluation might include hospital emergency department care, which can lead to lower admission rates for UTI patients. |
| **Avoidable Hospi-talization ACSC** | *Perforated appendix (#20)*<br><br>Lower rates may represent better quality of care. | • *Face Validity:* Perforation of appendix results from delay in surgery, potentially reflecting problems in access to prehospital and hospital care.<br>• *Precision:* Ruptured appendix is a common complication in patients hospitalized with appendicitis, suggesting that relatively precise estimates of area rate differences may be observed.<br>• *Minimum bias:* Ruptured appendix rates are higher in blacks, lower socioeconomic groups, the uninsured. Such population differences may | • *Precision:* Indicator demonstrates very substantial variation in performance across areas. However, almost all of that variation reflects unsystematic unobservable differences, rather than true differences in performance. Multivariate smoothing is particularly beneficial to this indicator.<br>• *Minimum bias:* Some of the observed variation in performance is due to systematic differences in patient characteristics. Area rankings and absolute performance are not affected by age-sex risk adjustment. | *HCUP Data:*<br>• Potentially an indirect measure of "quality," since area rates may not reflect quality of outpatient care.<br>• Much of the apparent differences in perforated appendix rates across areas are the result of random noise; use of "smoothed" estimates helps avoid such biases.<br>• Large variations across areas, suggesting notable opportunities for quality improvement. Avoidable hospitalizations are generally indicators of the quality of |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | contribute to observed differences in area rates, though evidence on this question is limited and the differences may reflect differences in access to care.<br>• *Construct validity:* Principal reason for occurrence of rupture appears to be delay in presentation to ER, though patient characteristics influencing decision to seek care and health system characteristics are also important.<br>• *Fosters true quality improvement:* Use of this quality indicator might lead to more performance of appendectomies in cases of questionable symptoms, in addition to reducing the occurrence of rupture.<br>• *Prior use:* Included in Weissman's set of avoidable hospitalizations and as a hospital-level measure in the previous version of HCUP I. | • *Construct validity:* Areas with higher rates of perforated appendix admissions also tend to have higher rates of admissions for other ACSCs. | outpatient care and other non-hospital aspects of the health care system, and so are likely to be of greatest value to public health authorities, planners, and purchasers. However, hospital practices may also have a (relatively minor) impact on perforated appendix rates.<br><br>*Further investigation:*<br>• Chart reviews and other supplemental data collection can distinguish possible causes of high rates, e.g., comorbid conditions, delay in presentation, delay in treatment. |
| **Avoidable Hospi-talization ACSC** | *Angina (#21)*<br><br>Lower rates may represent better quality of care. | • *Face Validity:* Effective treatments for coronary artery disease reduce admissions for serious complications of ischemic heart disease, including unstable angina.<br>• *Precision:* Angina is a very common cause of hospitalization, suggesting that reasonably precise estimates should be feasible.<br>• *Minimum bias:* Though genetic, behavioral, and socioeconomic risk factors for coronary artery disease may vary across areas, little evidence exists on the extent to which such differences account for area differences in rates. Thresholds for | • *Precision:* Indicator demonstrates substantial variation in performance across areas. Almost all of the observed variation reflects true differences in admission rates.<br>• *Minimum bias:* Area rankings are not affected by age-sex risk adjustment. However, risk adjustment does change the absolute performance of areas somewhat.<br>• *Construct validity:* Areas with high rates of angina admissions tend to have higher rates of other ACSC admissions. | *HCUP Data:*<br>• Potentially an indirect measure of "quality," since area rates may not reflect quality of outpatient care.<br>• Notable variations across areas, suggesting opportunities for quality improvement in care of ischemic heart disease patients.<br>• Accounting for measurable risk differences across areas affects absolute area performance.<br>• Consider in conjunction with other ACSC indictors of short-term complications of poor outpatient care, since all are generally positively correlated. |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | inpatient admission of angina patients (instead of management in the emergency department or other settings) may also differ across areas, though again little evidence exists on the importance of such practice differences.<br>• *Construct validity:* Little specific evidence on the validity of angina as an ACSC indicator exists.<br>• *Fosters true quality improvement:* Use of this quality indicator might raise thresholds for admission of angina patients, but no evidence exists on whether such practice changes would have adverse consequences.<br>• *Prior use:* Developed by Billings. | | • Interventions to improve outpatient care for ischemic heart disease can be targeted to patients treated by hospitals that admit relatively large shares of angina patients in an area.<br>*Further investigation:*<br>• Little previous literature to guide investigations of underlying causes of variations in angina admission rates.<br>• Examination of outpatient angina care may provide a more complete picture of care for this condition. This includes hospital emergency department care, which can lead to lower admission rates for patients with angina.<br>• Inpatient admissions plus emergency room visits for angina may provide a more complete picture of angina complication rates in an area. |
| **Avoidable Hospi-talization ACSC** | *Adult asthma (#22)*<br><br>Lower rates may represent better quality of care. | • *Face Validity:* One of the most common reasons for hospital admission. Consensus that asthma is readily treatable in an outpatient setting, using education and proper medication.<br>• *Precision:* Asthma is a common cause of admission and adults, and is likely to be measured somewhat precisely.<br>• *Minimum bias:* Environmental exposures to irritants such as allergens, tobacco smoke, and air pollution and occupation exposure to irritants may vary systematically by area and may affect admission rates. Admission rates vary by race, though | • *Precision:* Indicator demonstrates adequate variation in performance across areas. Most of the observed variation reflects true differences in admission rates.<br>• *Minimum bias:* Area rankings and absolute performance are not affected by age-sex risk adjustment.<br>• *Construct validity:* Areas with higher rates of adult asthma admission also tend to have higher rates of admission for the other ACSC conditions. | *HCUP Data:*<br>• Potentially an indirect measure of "quality," since area rates may not reflect quality of outpatient care.<br>• Notable variations across areas, suggesting opportunities for quality improvement in care of ischemic heart disease patients.<br>• Accounting for measurable risk differences across areas somewhat affects absolute performance.<br>• Consider in conjunction with other ACSC indictors of short-term complications of poor outpatient care, since all are generally positively correlated. |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | relationship with quality is not clear.<br>• *Construct validity:* Asthma hospitalization rates have been shown to vary with socioeconomic factors, insurance, self-rated area access to care, and to some extent inadequate care.<br>• *Fosters true quality improvement:* Care may shift to an outpatient setting, though there is not evidence suggesting the impact on patient outcomes.<br>• *Prior use:* Asthma admission rates is a HP 2010 goal (7.7 per 10,000 pop). The indicator was independently developed by both Billings and Weissman and is used by the UK National Health Service. | | • To track HP 2010 goal use together with pediatric asthma indicator.<br>• Interventions to improve outpatient care for asthma can be targeted to patients treated by hospitals that admit relatively large shares of angina patients in an area.<br>*Further investigation:*<br>• Environmental exposures may explain some of the differences in admission rates for adult asthma.<br>• Investigation of outpatient care, including compliance with guidelines may illuminate causes of high admission rates. This includes hospital emergency department care, which can lead to lower admission rates for patients with asthma.<br>• Inpatient admissions plus emergency room visits for asthma may provide a more complete picture of asthma exacerbation rates in an area. |
| **Avoidable Hospi-talization ACSC** | *Chronic obstructive pulmonary disease (COPD) (#23)*<br><br>Lower rates may represent better quality of care. | • *Face Validity:* COPD admissions are common and have been increasing. Based on evidence of effective treatments, guidelines have been developed to reduce COPD complications, including those leading to hospital admissions.<br>• *Precision:* COPD hospitalizations are very common, suggesting that relatively precise estimation of area differences is feasible.<br>• *Minimum bias:* Many patient factors influence progression and exacerbation of disease, including | • *Precision:* Indicator demonstrates substantial variation in performance across areas. Almost all of the observed variation reflects true differences in admission rates.<br>• *Minimum bias:* Some of the observed variation in performance is due to systematic differences in patient characteristics. Area rankings and absolute performance are somewhat affected by age-sex risk adjustment.<br>• *Construct validity:* Areas with higher COPD admission rates tend to have higher admission rates for other | *HCUP Data:*<br>• Potentially an indirect measure of "quality," since area rates may not reflect quality of outpatient care.<br>• Large variations across areas that are not due to random chance, suggesting significant opportunities for quality improvement in COPD care.<br>• Accounting for measurable risk differences across areas somewhat affects relative area rankings.<br>• Consider in conjunction with other ACSC indictors of short-term |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | smoking, socio-economic status and environmental exposures. Some miscoding of acute bronchitis as chronic bronchitis also occurs. Little evidence exists on the extent to which these factors account for area variations in COPD admission rates, though socioeconomic differences appear to be relatively unimportant.<br>• *Construct validity:* Up to 30% variance in admission rates has been associated with access to care. Though physicians often prescribe appropriate treatments for COPD, patient compliance with treatment is modest. Some evidence of underprovision of smoking cessation counseling and other advice related to behavior modifications that could reduce COPD hospitalizations, including low rates of counseling and discharge planning among patients hospitalized for COPD complications.<br>• *Fosters true quality improvement:* Some evidence that greater access to care leads to increased admissions, at least in the short run.<br>• *Prior use:* Developed by Billings. | ACSCs. | ACSC indictors of short-term complications of poor outpatient care, since all are generally positively correlated.<br>• Interventions to improve outpatient care can be targeted to COPD patients treated by hospitals that admit relatively large shares of patients in an area.<br>*Further investigation:*<br>• Evidence on area environmental exposures and smoking rates may help explain a substantial part of area variations in rates.<br>• Some evidence that many patients with COPD exacerbations are not counseled effectively in the hospital, so that HCUP data can be used to identify hospitals that may be able to improve counseling and post-discharge treatment for COPD patients.<br>• Review of COPD records may help identify specific causes of COPD admissions, so that patient clinical factors can be differentiated from environmental factors and problems with access to high-quality care.<br>• Examination of outpatient COPD care may provide a more complete picture of care for this condition, and better outpatient treatment may at least in the short run be associated with higher inpatient admission rates for COPD complications due to case-finding. |
| **Avoidable Hospi-** | *Congestive heart failure (CHF)* | • *Face Validity:* Considerable evidence shows that effective treatments for | • *Precision:* Indicator demonstrates substantial variation in performance | *HCUP Data:*<br>• Potentially an indirect measure of |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| **talization ACSC** | *(#24)*<br><br>Lower rates may represent better quality of care. | CHF can reduce morbidity and mortality, including hospitalizations. However, these treatments are often underprescribed, suggesting that differences in the quality of outpatient care for CHF may lead to significant differences in rates of hospitalization with complications.<br>• *Precision:* CHF is one of the most common causes of hospitalization; thus, reasonably precise estimates of area rates should be feasible.<br>• *Minimum bias:* Patient age, as well as a range of clinical measures of disease severity and comorbidity, may affect CHF complication rates. Limited evidence exists on the extent to which differences in these factors rather than differences in management strategies can explain area differences in CHF admission rates. Physicians in different areas may also differ in thresholds for inpatient admission for CHF, potentially biasing comparisons of complication rates across areas.<br>• *Construct validity:* Some evidence suggests that differences in insurance rates and in access to care may account for as much as 50% of the variation in CHF hospitalization rates.<br>• *Fosters true quality improvement:* Limited evidence suggests that incentives to encourage adherence to CHF treatment guidelines and effective outpatient management lead to lower CHF admission rates. This may occur because of changes in thresholds for admission (i.e., more | across areas. Almost all of the observed variation reflects true differences in admission rates.<br>• *Minimum bias:* Some of the observed variation in performance is due to systematic differences in patient characteristics. Area rankings and absolute performance are somewhat affected by age-sex risk adjustment. In particular risk adjustment impacts areas with very high or low performance.<br>• *Construct validity:* Areas with higher CHF admission rates tend to have higher admission rates for other ACSCs. | "quality," since area rates may not reflect quality of outpatient care.<br>• Large variations across areas, suggesting significant opportunities for quality improvement in CHF care.<br>• Accounting for measurable risk differences across areas somewhat affects relative area rankings.<br>• Interventions to improve outpatient care for CHF can be targeted to patients treated by hospitals that admit relatively large shares of CHF patients in an area.<br>*Further investigation:*<br>• Evidence on socioeconomic status (which can be proxied using patient residence zip code on discharge data) and insurance status of CHF patients may help explain area variations, though this relationship is complex, and may also reflect inadequate care.<br>• Review of CHF patient records may help identify specific causes of CHF admissions, so that patient clinical factors (including access to care and understanding of treatment needs) can be differentiated from environmental factors and problems with access to high-quality care.<br>• Examination of outpatient CHF care may provide a more complete picture of care for this condition. This includes hospital emergency department care, which can lead to lower admission rates for patients with CHF. |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | emergency room management) rather than true differences in complication rates.<br>• *Prior use:* Developed independently by Weissman and Billings. Used in DEMPAQ and UK National Health Service High Level Performance Indicators. | | • Inpatient admissions plus emergency room visits for CHF may provide a more complete picture of CHF complication rates in an area. |
| **Avoidable Hospi-talization ACSC** | *Diabetes short term complication (#25)*<br><br>Lower rates may represent better quality of care. | • *Face Validity:* High-quality outpatient management of diabetic patients has been shown to lead to reductions in almost all types of serious avoidable hospitalizations. However, tight control leads to more episodes of hypoglycemia, and may lead to more admissions as a result.<br>• *Precision:* Though admissions for short-term diabetic complications are not rare, little evidence exists on the extent to which rates vary systematically across areas, rather than due to random noise.<br>• *Minimum bias:* Minorities have higher rates of diabetes, and higher hospitalization rates may result in areas with higher minority concentrations. Disease-related risk factors for short term complications seem unlikely to vary by area. Areas that use more outpatient diabetes treatment may appear to have higher rates due to smaller denominator for this indicator. However, importance of this potential source of bias is not known.<br>• *Construct validity:* Patient non-compliance appears to be a significant causative factor in admissions. Some | • *Precision:* Indicator demonstrates adequate variation in performance across areas. Some of the observed variation reflects true differences in admission rates, though some reflects unsystematic variation.<br>• *Minimum bias:* Area rankings and absolute performance are not affected by age-sex risk adjustment.<br>• *Construct validity:* Areas with higher rates of diabetes short-term complication admissions also tend to have higher rates of admissions for other ACSCs. | *HCUP Data:*<br>• Potentially an indirect measure of "quality," since area rates may not reflect quality of outpatient care.<br>• Large variations across areas, suggesting significant opportunities for quality improvement in outpatient and preventive care. However, some variations may reflect biases and not true quality differences.<br>• Accounting for measurable risk differences across areas somewhat affects relative area rankings, though the impact is not substantial.<br>• Consider in conjunction with other ACSC indictors, especially other indicators of diabetes care.<br>• Interventions to improve outpatient care can be targeted to patients treated by hospitals that admit relatively large shares of short-term complications.<br>*Further investigation:*<br>• Examination of outpatient diabetes care may provide a more complete picture of diabetes care.<br>• Although not demonstrated conclusively, differing population rates of diabetes may impact this |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | non-compliance has been reported as due to access to care problems. Other relationships between admission rates and access to care have been noted. Some admissions are in new cases, suggesting problems in outpatient detection.<br>• *Fosters true quality improvement:* There is no evidence that this indicator is likely to result in perverse incentives.<br>• *Prior use:* Included in Weissman, Billings, DEMPAQ, and the UK National Health Service High Level Indicator. Also a hospital-level measure in the previous version of HCUP I. | | indicator. Therefore, areas may wish to consider performance in conjunction with these rates. |
| **Avoidable Hospi-talization ACSC** | *Uncontrolled diabetes (#26)*<br><br>Lower rates may represent better quality of care. | • *See evidence for short-term diabetes complications.* | • *Precision:* Indicator demonstrates adequate variation in performance across areas. Most of the observed variation reflects true differences in admission rates.<br>• *Minimum bias:* Area rankings and absolute performance are not affected by age-sex risk adjustment.<br>• *Construct validity:* Areas with higher rates of diabetes short-term complication admissions also tend to have higher rates of admissions for other ACSCs. | *HCUP Data:*<br>• Potentially an indirect measure of "quality," since area rates may not reflect quality of outpatient care.<br>• Large variations across areas, suggesting significant opportunities for quality improvement in outpatient and preventive care. However, some variations may reflect biases and not true quality differences.<br>• Accounting for measurable risk differences across areas somewhat affects relative area rankings, particularly in areas with higher rates.<br>• Consider in conjunction with other ACSC indictors, especially other indicators of diabetes care. To track HP2010 goal, use in conjunction with diabetes short-term |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | | | complications indicator.<br>• Interventions to improve outpatient care can be targeted to patients treated by hospitals that admit relatively large shares of short-term complications.<br>*Further investigation:*<br>• Examination of outpatient diabetes care may provide a more complete picture of diabetes care.<br>• Although not demonstrated conclusively, differing population rates of diabetes may impact this indicator. Therefore, areas may wish to consider performance in conjunction with these rates. |
| **Avoidable Hospi-talization ACSC** | *Diabetes long-term complications (#27)*<br><br>Lower rates may represent better quality of care. | • *Face Validity:* Substantial evidence that better glycemic control reduces long-term complications, and substantial evidence of failure to comply with treatment guidelines for glycemic control. Long-term diabetic complications are common in the US population.<br>• *Precision:* Though hospital complications related to diabetes are not uncommon, little evidence exists on the extent to which complication admissions vary across areas.<br>• *Minimum bias:* Complication rates are higher in minority populations, and may also vary with socioeconomic status. Whether such differences in area populations explains variations in long-term complication rates is unclear. Some evidence suggests that discharge data has relatively low predictive power for true diabetic | • *Precision:* Indicator demonstrates adequate variation in performance across areas. Most of the observed variation reflects true differences in admission rates.<br>• *Minimum bias:* Area rankings are substantially affected by age-sex risk adjustment, suggesting that patient characteristics influence performance on indicator. However, absolute performance is only moderately impacted.<br>• *Construct validity:* Areas with higher rates of diabetes long-term complication admissions also tend to have higher rates of admissions for other ACSCs. | *HCUP Data:*<br>• Potentially an indirect measure of "quality," since area rates may not reflect quality of outpatient care.<br>• Large variations across areas, suggesting significant opportunities for quality improvement in outpatient and preventive care.<br>• Accounting for measurable risk differences across areas affects relative area rankings, particularly in areas with higher rates.<br>• Because long-term complications may take years to develop, this indicator may capture different aspects of care than other ACSC indicators.<br>• Consider in conjunction with other ACSC indictors, especially other indicators of diabetes care.<br>• Interventions to improve outpatient diabetic care to reduce long-term |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | complications, but the implications for area rate comparisons are unclear. *Construct validity:* Though access to good diabetes care and patient compliance are clearly important in avoiding complications, little evidence exists on the extent to which differences in quality of outpatient care explain differences in long-term complication rates.<br>• *Fosters true quality improvement:* No evidence that this indicator would establish perverse incentives.<br>• *Prior use:* Included in the previous version of the HCUP I indicator set. | | diabetic care to reduce long-term complications can be targeted to patients by hospitals that admit large shares of the area patients with complications.<br>*Further investigation:*<br>• Diabetes long-term complications may also be diagnosed and treated in an outpatient setting. Examination of outpatient diabetes care may provide a more complete picture of diabetes care.<br>• Collection of supplemental clinical data on patients with long-term complications may provide further insights into the underlying causes, but relatively little literature exists to guide such analyses.<br>• Although not demonstrated conclusively, differing population rates of diabetes may impact this indicator. Therefore, areas may wish to consider performance in conjunction with these rates. |
| **Avoidable Hospi-talization ACSC** | *Hypertension (#28)*<br><br>Lower rates may represent better quality of care. | • *Face Validity:* Effective ambulatory treatments of hypertension exist, and adherence to established guidelines for hypertension treatments has been shown to reduce hypertension-related complications, including hospitalizations.<br>• *Precision:* Little evidence exists on the precision of area rates of hypertension complications. Hospitalizations for hypertension are relatively uncommon, so that estimates may not be precise.<br>• *Minimum bias:* Little evidence on | • *Precision:* Indicator demonstrates adequate variation in performance across areas. Some to most of the observed variation reflects true differences in admission rates.<br>• *Minimum bias:* Area rankings and absolute performance are not affected by age-sex risk adjustment.<br>• *Construct validity:* Areas with higher rates of hypertension admissions also tend to have higher rates of admissions for other ACSCs. | *HCUP Data:*<br>• Potentially an indirect measure of "quality," since area rates may not reflect quality of outpatient care.<br>• Significant variations across areas that are not due to random chance, suggesting opportunities for quality improvement.<br>• Accounting for measurable risk differences across areas somewhat affects area absolute performance.<br>• Interventions to improve outpatient care can be targeted to hypertension patients treated by hospitals that |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | potential biases associated with this indicator exists, though one study found a significant impact of age and sex adjustment.<br>• *Construct validity:* Access to care problems associated with relatively high area admission rates in one study, but evidence is limited.<br>• *Fosters true quality improvement:* Little evidence on the effects of use of this indicator.<br>• *Prior use:* Developed independently by Billings and Weissman. | | admit relatively large shares of patients with complications in an area.<br>*Further investigation:*<br>• Limited evidence on the relative importance of socioeconomic factors, comorbid diseases, and access to high-quality medical care on complication rates. |
| **Avoidable Hospi-talization ACSC** | *Lower extremity amputation in diabetics (#29)*<br><br>Lower rates may represent better quality of care. | • *Face Validity:* Lower extremity amputation affects up to 15% of diabetics and diabetes is the leading cause of LE amputation. Many of the potential causes of disease leading to LE amputation are preventable with proper foot care and tight glucose control. There is evidence that recommendations for the above preventative measures are not widely adhered to.<br>• *Precision:* Incidence varies widely internationally, and is common enough to expect adequate precision.<br>• *Minimum bias:* Age and sex may affect rate and may vary systematically by area. LE rates also vary by race, though the relationship with quality of care is unknown.<br>• *Construct validity:* Tight glucose control and education programs have been shown to reduce risk factor and/or amputation rates directly, though few studies had control groups.<br>• *Fosters true quality improvement:* | • *Precision:* Indicator demonstrates adequate variation in performance across areas. After smoothing, most of the observed variation reflects true differences in admission rates. Multivariate smoothing is particularly beneficial to this indicator.<br>• *Minimum bias:* Area rankings and absolute performance are somewhat affected by age-sex risk adjustment.<br>• *Construct validity:* Lower-extremity amputation appears to be somewhat independent from the other ACSC indicators. | *HCUP Data:*<br>• Potentially an indirect measure of "quality," since area rates may not reflect quality of outpatient care.<br>• Though admission rates vary adequately across areas, some of the variation is due to unsystematic variation; "smoothed" estimates help avoid this problem.<br>• Accounting for measurable risk differences across areas somewhat affects area performance.<br>• Large variations across areas, suggesting significant opportunities for quality improvement in outpatient and preventive care.<br>• Because long-term complications may take years to develop, this indicator may capture different aspects of care than other ACSC indicators.<br>• Consider in conjunction with other ACSC indictors, especially other indicators of diabetes care. |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | Given the severity of conditions requiring LE amputation, it is unlikely care would be denied.<br>• *Prior use:* Healthy People 2010 goal (1.8 per 1000 persons with diabetes). Included in the DEMPAQ measure set for outpatient care. | | *Further investigation:*<br>• Although not demonstrated conclusively, differing population rates of diabetes may impact this indicator. Therefore, areas may wish to consider performance in conjunction with these rates. |
| **Avoidable Hospi-talization ACSC** | *Low birth weight (#30)*<br><br>Lower rates may represent better quality of care. | • *Face Validity:* Better prenatal care and education may reduce the rate of low birth weight births, though effects may only be significant in high-risk pregnancies.<br>• *Precision:* Although low birth weight births are relatively infrequent, the large number of births in most areas suggests that this indicator can be precisely measured.<br>• *Minimum bias:* Age, race, socioeconomic status, behaviors such as smoking and substance abuse, and many other factors influence the risk of low weight births, and may be indirectly related to access to high-quality prenatal care. Little evidence exists on the extent to which each of these factors in addition to use of prenatal care contribute to differences in low birth weight rates across areas.<br>• *Construct validity:* While a number of studies have documented effects of particular interventions on the rate of low birth weight births, particularly in high-risk populations, less evidence exists on the effects of better prenatal care in an area on area low birth weight rates.<br>• *Fosters true quality improvement:* No | • *Precision:* Indicator demonstrates very substantial variation in performance across areas. After smoothing, most of the observed variation reflects true differences in admission rates. Multivariate smoothing is particularly beneficial to this indicator.<br>• *Minimum bias:* No appropriate risk adjustment available (only infant gender is available).<br>• *Construct validity:* Low birth weight appears to be independent of the other ACSC indicators. | *HCUP Data:*<br>• Potentially an indirect measure of "quality," since area rates may not reflect quality of outpatient care.<br>• Though admission rates vary substantially across areas, some of the variation is due to unsystematic variation; "smoothed" estimates help avoid this problem.<br>• Interventions to improve quality of prenatal care can be targeted to hospitals that admit relatively large shares of low birth weight births in an area.<br>*Further investigation:*<br>• Socioeconomic, behavioral, and other factors that are only indirectly related to prenatal care appear to have substantial influences on low birth weight rates. Though socioeconomic measures typically are not included in hospital discharge records, information on patient insurance status and zip of residence can be used to develop useful proxies for socioeconomic status.<br>• Review of low birth weight birth records may help identify specific causes of the low birth weight |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | significant problems relating to the use of the indicator have been identified, beyond the fact that many factors other than medical care influence low birth weight rates.<br>• *Prior use:* Included in the previous version of the HCUP I indicator set and in HEDIS. Used by University Hospital Consortium, and United Health Care. | | complications, so that patient clinical factors can be differentiated from environmental factors and problems with access to high-quality care.<br>• Additional linkage to maternal records may also provide additional means for risk adjustment.<br>• Direct examination of prenatal care use in an area may provide a more complete picture of the reasons for differences in low birth weight rates. |
| **Avoidable Hospi-talization ACSC** | *Pediatric asthma (#31)*<br><br>Lower rates may represent better quality of care. | • *Face Validity:* Asthma is the most common chronic childhood disease. Many treatments have been shown to reduce pediatric asthma complications, including those leading to hospitalization.<br>• *Precision:* Because hospitalizations for pediatric asthma are very common, relatively precise estimates of area rates should be feasible. Previous studies have documented substantial variation in hospitalization rates across areas.<br>• *Minimum bias:* Some children may be at increased risk for admission due to exposure to environmental triggers, comorbidities, and genetic factors, though the extent to which these factors can explain area differences is not known. Socioeconomic differences appear to account for a large share of area variations in asthma hospitalization rates.<br>• *Construct validity:* Adherence to guidelines for treatments demonstrated | • *Precision:* Indicator demonstrates substantial variation in performance across areas. Most of the observed variation reflects true differences in admission rates.<br>• *Minimum bias:* Area rankings and absolute performance are not affected by age-sex risk adjustment.<br>• *Construct validity:* Areas with higher rates of pediatric asthma admission also have higher rates of admission for other ACSCs. | *HCUP Data:*<br>• Potentially an indirect measure of "quality," since area rates may not reflect quality of outpatient care.<br>• Large variations across areas, suggesting significant opportunities for quality improvement in outpatient asthma care.<br>• Accounting for measurable risk differences across areas somewhat affects area absolute performance.<br>• Consider in conjunction with other ACSC indictors of short-term complications of poor outpatient care, since all are generally positively correlated.<br>• Interventions to improve outpatient care can be targeted to asthma patients treated by hospitals that admit relatively large shares of asthma patients in an area.<br>*Further investigation:*<br>• Socioeconomic associations with asthma hospitalization rates appear to be very strong. Though |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | to be effective in asthma care has been shown to lower hospitalization rates, and physician expert panels have also concluded that effective treatment lowers hospitalization rates.<br>• *Fosters true quality improvement:* Though it is possible that use of this indicator may lead to reduced appropriate hospitalization for asthma, there is no evidence suggesting that use of the indicator worsens outcomes of care.<br>• *Prior use:* Included in South Carolina's adaptation of Billings, and as a provider measure in the previous HCUP I indicator set. Similar measure used by United Health Care. | | socioeconomic measures typically are not included in hospital discharge records, information on patient insurance status and zip of residence can be used to develop useful proxies for socioeconomic status.<br>• Review of asthma patient records may also help identify specific causes of asthma complications, so that patient clinical factors can be differentiated from environmental factors and problems with access to high-quality care.<br>• Examination of outpatient asthma care may provide a more complete picture of the reasons for differences in asthma admission rates. |
| **Avoidable Hospi-talization ACSC** | *Pediatric gastroenteritis (#32)*<br><br>Lower rates may represent better quality of care. | • *Face Validity:* Guidelines have been established for outpatient management of pediatric gastroenteritis, and adherence rates are relatively low. Experts have concluded that greater adherence would reduce hospitalization rates.<br>• *Precision:* Hospitalizations for pediatric gastroenteritis are very common, and hospitalization rates vary widely, suggesting that true differences across areas can be determined reliably.<br>• *Minimum bias:* Admission rates vary by socio-economic status, parental coping ability, and clinical comorbidities. Relative importance of these factors is largely unknown, as is | • *Precision:* Indicator demonstrates substantial variation in performance across areas. Most of the observed variation reflects true differences in admission rates.<br>• *Minimum bias:* Area rankings and absolute performance are not affected by age-sex risk adjustment.<br>• *Construct validity:* Areas with higher admission rates for pediatric gastroenteritis tend to have higher admission rates for other ACSCs. | *HCUP Data:*<br>• Potentially an indirect measure of "quality," since area rates may not reflect quality of outpatient care.<br>• Large variations across areas that are not due to random chance, suggesting significant opportunities for quality improvement.<br>• Accounting for measurable risk differences across areas somewhat affects absolute area performance.<br>• Consider in conjunction with other ACSC indictors of short-term complications of poor outpatient care, since all are generally positively correlated.<br>• Interventions to improve outpatient care can be targeted to |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | the extent to which socioeconomic differences are due to correlated differences in access to high-quality care versus other factors.<br>• *Construct validity:* The relationship of adherence to guidelines and quality or increased admissions has not been established. High rate of apparently "inappropriate" admissions has been documented. Parental compliance with prescribed treatment is low, suggesting that problems with patient education may be important.<br>• *Fosters true quality improvement:* Some evidence suggests that expanded access to care (e.g., Medicaid expansions) may not lead to net reductions in hospitalizations for gastroenteritis.<br>• *Prior use:* Included in South Carolina's adaptation of Billings. | | gastroenteritis patients treated by hospitals that admit relatively large shares of patients in an area.<br>*Further investigation:*<br>• Socioeconomic differences appear moderately important in explaining area rate differences. Patient zip code of residence can be used as a proxy for socioeconomic status in investigating this question, and supplemental data collection can help determine whether medical or other factors are responsible for the socioeconomic differences.<br>• Review of patient records may help identify specific causes of gastroenteritis admissions, so that patient clinical factors can be differentiated from environmental factors and problems with access to high-quality care.<br>• Examination of outpatient gastroenteritis care may provide a more complete picture of the reasons for high or low rates of admissions with complications. Better outpatient treatment may not reduce admission rates for gastroenteritis, perhaps due to greater case-finding. |

**Table 13. Summary evidence table for mortality indicators**

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| **In-Hospital Mortality for Inpatient Conditions** | *AMI mortality (#33)*<br><br>Lower AMI mortality may represent better quality. | • *Face validity:* Many acute treatments have been shown to reduce short- and long-term AMI mortality; many effective AMI treatments are known to be underused, suggesting important opportunities for quality improvement.<br>• *Precision:* Mortality outcomes vary substantially across hospitals even after detailed risk adjustment. However, because many hospitals treat a relatively small number of AMI patients, much of this apparent variation is likely to be noise.<br>• *Minimum bias:* Some variation in AMI patient mortality risk exists across hospitals; because some of this risk is related to AMI severity, it may be difficult to adjust out using discharge data alone. However, because AMI is an urgent medical condition, AMI patients differ less across hospitals than patients with more elective conditions (e.g., CABG patients). Thus, "selection" bias may not be substantial. Differences in hospital practices related to transfers and length of stay may also lead to biased measurement of true mortality differences across hospitals.<br>• *Construct validity:* Treatments have been shown to reduce patient level mortality as well as provider level mortality rates. Hospitals identified as providing quality treatment for AMIs (as measured by processes of care) also tend to have lower | • *Precision:* Indicator demonstrates substantial variation in performance across providers. Some of the observed variation reflects true differences in mortality , though some reflects unsystematic variation. Multivariate smoothing is particularly beneficial to this indicator.<br>• *Minimum bias:* Some of the observed variation in performance is due to systematic differences in patient characteristics. Provider rankings are substantially impacted by APR-DRG risk adjustment, suggesting that patient characteristics influence performance on indicator. In particular risk adjustment impacts providers with very high or low performance.<br>• *Construct validity:* Though AMI mortality does not load strongly on any factors, it is positively correlated with in-hospital mortality for a number of other conditions and procedures, including CABG. | *HCUP Data Use:*<br>• Consider in conjunction with other QI measures related to in-hospital mortality and cardiac care.<br>• Though AMI mortality varies substantially across providers, most of the variation is due to unsystematic variation; "smoothed" estimates help avoid this problem. Important to remove random variation, through MSX smoothing or other techniques, in mortality outcomes, as many hospitals treat few AMI patients.<br>• Risk adjustment with APR-DRG at minimum is recommended for this indicator, due to evidence of potential bias.<br><br>*Further Investigation:*<br>• If possible, track mortality after hospitalization, to avoid biases in apparent survival due to differences in hospital discharge and transfer practices. Consider hospital average length of stay and transfer rates.<br>• Medical chart reviews can be used to determine whether hospitals with better performance have less severe patients, and have better adherence to AMI treatment guidelines and practices known to reduce mortality. |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | mortality rates, though these findings are contradicted in a few studies. Such contradictions may be due to general insensitivities of mortality rates to process measures. Some states have achieved lower mortality rates concurrently with improving performance on process measures of quality.<br>• Fosters quality improvement: Since providers tend to have little choice on which patients to treat for AMI, bias due to selection is unlikely. However, it is possible that this measure would encourage differences in admission and length of stay, such that deaths would not show up on hospital discharge data.<br>• *Prior use*: AMI mortality is one of the most commonly used indicators. It is currently used by IMSystem, Collaborative Approach to Resource Effectiveness (CARE) Project, HealthGrades.com, Greater New York hospital Association, University Hospital Consortium, Pennsylvania Health Care Cost Containment Council, UK National Health Service High Level Performance Indicators, and California Hospital Outcomes Project. The following use this indicator under certain conditions: Michigan Hospital Association (in aggregated measure) and Cleveland Health Quality Choice (grouped with cardiovascular measure). The use of this indicator has been proposed by Colorado Health and Hospital. This | | |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | indicator has also been designated as a JCAHO core measure. | | |
| **In-Hospital Mortality for Inpatient Conditions** | *Congestive Heart Failure (CHF) (#34)*<br><br>Lower CHF mortality may represent better quality. | • *Face validity*: Common condition, with relatively frequent inpatient mortality. Recent studies show a decline in case fatality rates and some improvement in population mortality rates, possibly related to a variety of recent innovations in the care of patients hospitalized CHF patients.<br>• *Precision*: CHF hospitalizations are common, and wide variations in mortality rates for hospitalized CHF patients have been reported. Both of these features suggest that significant mortality differences may be identifiable.<br>• *Minimum bias*: Patient disease severity (e.g., extent of symptoms, stage of heart failure) and comorbidities (e.g., diabetes, chronic respiratory disease) have substantial effects on inpatient mortality. But little evidence on whether discharge-based or clinically-detailed risk adjustment methods significantly affect measured hospital performance. Because a primary diagnosis of CHF is not very sensitive in administrative data (i.e., many patients with CHF are missed), it is possible that coding differences across hospitals will bias mortality comparisons; little evidence exists on the importance of this problem. Some CHF patients can be managed with outpatient treatment; because these patients tend to be less severely ill, hospitals that manage CHF more | • *Precision*: Indicator demonstrates adequate variation in performance across providers. Some of the observed variation reflects true differences in mortality, though some reflects unsystematic variation. Multivariate smoothing is particularly beneficial to this indicator.<br>• *Minimum bias*: Provider rankings are somewhat affected by APR-DRG risk adjustment. In particular risk adjustment impacts providers with very high or low performance.<br>• *Construct validity*: Performance is related to mortality for some other medical conditions (including stroke, pneumonia, GI bleed, and hip fracture). | *HCUP Data Use:*<br>• Consider in conjunction with other QI measures related to in-hospital mortality for other common conditions and procedures.<br>• Literature on heterogeneity of CHF patients suggests that differences in comorbidity and coding might affect measured performance; thus, caution in interpreting mortality differences is warranted.<br>• Though CHF mortality differs substantially across hospitals, most of the apparent differences reflect random noise; "smoothed" estimates avoid this problem.<br>• Risk adjustment for comorbidities reduces bias, and significantly affects relative hospital performance.<br><br>*Further Investigation:*<br>• If possible, very helpful to track mortality after hospitalization, to avoid biases in apparent hospital mortality rates due to differences in hospital discharge and transfer practices. Consider hospital average length of stay and transfer rates in interpreting measures.<br>• Because "best practice" at some hospitals may involve outpatient management of lower-risk cases, assessment of CHF patients managed as hospital outpatients may be valuable complement to analysis of inpatient mortality – in particular, do |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | effectively on an outpatient basis may have poorer measured inpatient performance.<br>• *Construct validity*: Hospitals differ widely in processes of care known to reduce CHF mortality (e.g., use of ACE inhibitors and echocardiographic evaluation).<br>• *Fosters true quality improvement*: Coding practices (e.g., whether CHF is listed as a primary or secondary diagnosis in patients with serious comorbid illnesses) may respond to use of CHF as a quality indicator. Might discourage outpatient management of less severe CHF cases, and encourage discharges or transfers of patients with severe CHF. | | hospitals with high inpatient mortality also have relatively high rates of outpatient (emergency department) management, suggesting they are admitting a more severely ill subset of the CHF patients they treat?<br>• Chart reviews may be helpful to determine whether apparent mortality differences are the result of differences in coding, comorbidities, and/or disease severity. |
| **In-Hospital Mortality for Inpatient Conditions** | *GI hemorrhage mortality (#35)*<br><br>Lower GI hemorrhage mortality may represent better quality. | • *Face validity:* GI hemorrhage mortality rates vary substantially, and a range of treatments (e.g., early endoscopy, vasopressin for many types of patients) have been shown to reduce mortality.<br>• *Precision:* Mortality rates for GI hemorrhage vary substantially across hospitals, and the condition is commonly treated at many hospitals.<br>• *Minimum bias:* The nature of the bleed and its severity, and especially comorbid diseases, may influence GI bleed mortality and may not be well-captured in hospital discharge data. However, GI bleed is an urgent medical condition, suggesting that opportunities for selection bias across hospitals may be limited. One study suggested that such differences would | • *Precision*: Indicator demonstrates adequate variation in performance across providers. Some of the observed variation reflects true differences in mortality, though some reflects unsystematic variation. Multivariate smoothing is particularly beneficial to this indicator.<br>• *Minimum bias*: Some of the observed variation in performance is due to systematic differences in patient characteristics. Provider rankings are somewhat affected by APR-DRG risk adjustment. In particular risk adjustment impacts providers with very high or low performance.<br>• *Construct validity*: Performance is related to mortality for some other medical conditions (including stroke, pneumonia, hip fracture, and CHF). | HCUP Data Use:<br>• Consider in conjunction with other QI measures related to in-hospital mortality for other common conditions and procedures.<br>• Though GI hemorrhage mortality differs substantially across hospitals, most of the apparent differences reflect random noise; "smoothed" estimates avoid this problem.<br>• Risk adjustment for APR-DRGs reduces bias, and significantly affects relative hospital performance.<br><br>Further Investigation:<br>• It may be helpful to track mortality after hospitalization, to avoid biases in apparent hospital mortality rates due to differences in hospital |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | alter hospital performance rankings in many cases; however, this study did not explicitly remove random noise.<br>• *Construct validity:* Many processes of care lead to better bleeding control and lower risk of re-bleeding, although there is little evidence of mortality differences. In many cases, GI hemorrhage deaths may be related to comorbid conditions, although it is possible that the effectiveness of management of GI bleeding in the context of severe underlying diseases may affect mortality from these comorbidities. Construct validity at the provider level not established.<br>• *Fosters true quality improvement:* Possible that use of GI hemorrhage mortality as quality indicator would lead to earlier discharges or transfers, improving measured performance without actually improving outcomes.<br>• *Prior use:* GI hemorrhage mortality is discussed in the literature as an indicator and outcome measure. It is used by Michigan Hospital Association (in aggregated measure), Cleveland Health Quality Choice, and Maryland QI Project. | | discharge and transfer practices. Consider hospital average length of stay and transfer rates in interpreting measures.<br>• Because "best practice" at many hospitals may involve early endoscopy and outpatient management of low-risk bleeds, assessment of GI bleed patients managed as hospital outpatients may be valuable complement to analysis of inpatient mortality – in particular, do hospitals with high inpatient mortality have relatively high rates of outpatient management? |
| **In-Hospital Mortality for Inpatient Conditions** | *Hip fracture mortality (#36)*<br><br>Lower hip fracture mortality may represent better quality. | • *Face validity:* Hip fracture is associated with a heightened risk of death, because fractures typically occur in relatively frail patients at risk for deep venous thrombosis, pneumonia, and myocardial ischemia.<br>• *Precision:* Relatively common condition and outcome – most common type of fracture requiring | • *Precision:* Indicator demonstrates substantial variation in performance across providers. Some of the observed variation reflects true differences in mortality, though some reflects unsystematic variation. Multivariate smoothing is particularly beneficial to this indicator.<br>• *Minimum bias*: Provider rankings are | HCUP Data Use:<br>• Consider in conjunction with other QI measures related to in-hospital mortality for hip replacement.<br>• Though hip fracture mortality differs substantially across hospitals, some of the apparent differences reflect random noise; "smoothed" estimates avoid this problem. |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | hospitalization, with in-hospital mortality approaching 5%. Suggests somewhat precise estimates at the provider level are feasible.<br>• *Minimum bias:* Although all patients with known hip fractures are hospitalized, some selection bias may occur as a result of certain hospitals treating more complex or frail patients. Patient demographics and comorbidities are important predictors of short-term mortality that can be obtained from discharge data; major functional impairment (ADL) is also an important predictor, but requires prospective data collection. Little evidence on whether these additional patient factors differ substantially across hospitals.<br>• *Construct validity:* Better adherence to processes of care that prevent hip fracture complications (e.g., DVT prophylaxis, beta blockers, state-of-the-art surgical techniques) may be associated with lower mortality at the patient level. Construct validity at the provider level not established.<br>• *Fosters true quality improvement:* Possible that use of hip fracture mortality as QI would encourage earlier discharges, e.g. to nonacute care facilities, since 30-day mortality is considerably higher than inpatient mortality. Hip fracture mortality QI not likely to reduce access to surgery, because fracture treatment is not an elective procedure. Because comorbidities can significantly affect | somewhat affected by APR-DRG risk adjustment.<br>• *Construct validity:* Significantly related to in-hospital mortality for other medical conditions including GI hemorrhage, stroke, pneumonia, and CHF. | • Risk adjustment for APR-DRGs reduces bias, and significantly affects relative hospital performance, and is recommended<br><br>Further Investigation:<br>• If possible, very helpful to track mortality after hospitalization, to avoid biases in apparent hospital mortality rates due to differences in hospital discharge and transfer practices. Consider hospital average length of stay and transfer rates in interpreting measures.<br>• Because explicit evidence on practices that reduce short-term mortality after hip fracture is relatively limited, detailed chart reviews may be more helpful in improving QI use by identifying differences in functional status, severity, and other risk factors that are not well-measured in hospital discharge data but that may differ systematically across hospitals. |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | mortality, possible that hospitals could improve measured performance by "upcoding."<br>• *Prior use:* Use of hip fracture mortality as an indicator is discussed in the literature. It is used by UK National Health Service High Level Performance Indicators and University Hospital Consortium (DRG 209). | | |
| **In-Hospital Mortality for Inpatient Conditions** | *Pneumonia mortality (#37)*<br><br>Lower pneumonia mortality may represent better quality. | • *Face validity*: Hospitalizations for pneumonia are very common, in-hospital death is relatively common (around 75,000 per year). Although patient characteristics are a principal determinant of mortality, better processes of care may also improve patient outcomes.<br>• *Precision*: Substantial patient heterogeneity reduces precision of hospital estimates; previous studies suggest a relatively high probability of errors in labeling hospitals as high-mortality outliers in basic statistical comparisons.<br>• *Minimum bias*: Hospitals may differ in the extent to which they manage relatively uncomplicated pneumonia cases as outpatients. They also may differ in the extent to which they transfer "terminal" patients with pneumonia to home care or nonacute centers. Finally, hospitals may differ in how they code the nature and severity of pneumonia cases in discharge abstracts. Though some risk adjustment is possible with administrative data, additional clinical indicators have additional predictive | • *Precision*: Indicator demonstrates substantial variation in performance across providers. Most of the observed variation reflects true differences in mortality. Multivariate smoothing is particularly beneficial to this indicator.<br>• *Minimum bias*: Some of the observed variation in performance is due to systematic differences in patient characteristics. Provider rankings are substantially by APR-DRG risk adjustment, suggesting that patient characteristics influence performance on indicator. In particular risk adjustment impacts providers with very high or low performance.<br>• *Construct validity*: Performance is related to mortality for some other medical conditions (including stroke, CHF, hip fracture and GI bleed). | *HCUP Data Use:*<br>• Consider in conjunction with other QI measures related to in-hospital mortality for other common conditions and procedures.<br>• Literature on heterogeneity of pneumonia patients suggests that differences in comorbidity and coding might affect measured performance; thus, caution in interpreting mortality differences is warranted.<br>• Much of the apparent variation in pneumonia mortality rates across hospitals in basic statistical comparisons reflects random noise; "smoothed" estimates avoid this problem.<br>• Risk adjustment for comorbidities reduces bias, and affects relative hospital performance.<br><br>*Further Investigation:*<br>• If possible, very helpful to track mortality after hospitalization, to avoid biases in apparent hospital mortality rates due to differences in hospital discharge and transfer practices. Consider hospital average length of |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | capacity. All of these factors may bias hospital mortality comparisons, though little evidence exists on the magnitude of these potential biases.<br>• *Fosters true quality improvement*: Better adherence to certain processes of inpatient care (e.g., rapidity and choice of antibiotics) leads to better outcomes, including mortality. Use as a quality indicator may provide incentives to improve processes, though evidence to date only has shown a relationship between hospital structural characteristics and hospital-level mortality, not processes of care. Hospitals may also respond by changes in admission, coding, and discharge practices, rather than taking steps to improve patient mortality. | | stay and transfer rates in interpreting measures.<br>• Because "best practice" at some hospitals may involve outpatient management of lower-risk cases, assessment of pneumonia patients managed as hospital outpatients may be valuable complement to analysis of inpatient mortality – in particular, do hospitals with high inpatient mortality also have relatively high rates of outpatient (emergency department) management, suggesting they are admitting a more severely ill subset of the pneumonia patients they treat?<br>• Chart reviews may be helpful to determine whether apparent mortality differences are the result of differences in coding, comorbidities, and/or disease severity. In particular, analysis should consider differences in hospital practices for using ICD-9 codes that lead to a classification of DRG 79 (acute respiratory infection) versus DRG 89 (pneumonia). |
| **In-Hospital Mortality for Inpatient Conditions** | *Acute stroke mortality (#38)*<br><br>Lower acute stroke mortality may represent better quality. | • *Face validity*: Common cause of admission and in-hospital mortality, though acute stroke mortality is determined primarily by stroke severity and most deaths occur after the initial acute hospitalization.<br>• *Precision*: In-hospital mortality for stroke patients is somewhat common (10-15% of cases). However, because in-hospital mortality is related primarily to stroke severity, hospital mortality estimates may be subject to considerable random noise. | • *Precision*: Indicator demonstrates substantial variation in performance across providers. Most of the observed variation reflects true differences in mortality. Multivariate smoothing is particularly beneficial to this indicator.<br>• *Minimum bias*: Some of the observed variation in performance is due to systematic differences in patient characteristics. Provider rankings are somewhat affected by APR-DRG risk adjustment. | *HCUP Data Use:*<br>• Consider in conjunction with other QI measures related to in-hospital mortality for other common conditions and procedures.<br>• Because positive predictive value of stroke diagnosis codes is suboptimal, and because some strokes will be missed using standard diagnosis codes, caution in interpreting mortality differences is warranted.<br>• Much of the apparent variation in stroke mortality rates across hospitals |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | • *Minimum bias*: Diagnosis codes for stroke have some sensitivity and specificity problems, with a positive predictive value of less than 80% for new stroke associated with the diagnosis codes included in the stroke DRG. Hospitals may also differ in their admission practices for relatively uncomplicated strokes and transient ischemic attacks (TIAs). Because most stroke deaths occur many days to weeks after the stroke, differences in hospital practices for discharging or transferring patients with poor prognoses may also affect in-hospital mortality. Some important predictors of in-hospital mortality can be obtained from discharge data. Previous studies reported sensitivity of hospital ranking to the risk adjustment method used; however, these studies did not account for random noise that might have contributed to differences in risk adjustment effects. All of these factors may bias hospital mortality comparisons, though little evidence exists on the magnitude of these potential biases.<br>• *Construct validity*: Hospital treatments such as rapid intervention (possibly including thrombolysis) and use of specialized stroke units have some impact on acute mortality.<br>• *Fosters true quality improvement*: In addition to improving quality of care, hospitals may respond to this quality indicator by changes in admission, coding, and discharge practices. | • *Construct validity*: Performance is related to mortality for some other medical conditions (including pneumonia, hip fracture, CHF, and GI bleed). | in basic statistical comparisons reflects random noise; "smoothed" estimates avoid this problem.<br>• Risk adjustment for comorbidities reduces bias, and affects relative hospital performance.<br>*Further Investigation:*<br>• If possible, very helpful to track mortality after hospitalization, to avoid biases in apparent hospital mortality rates due to differences in hospital discharge and transfer practices. Consider hospital average length of stay and transfer rates in interpreting measures.<br>• Because "best practice" at some hospitals may involve failure to admit some low-risk cases, exploration of the extent to which stroke patients are treated as outpatients using emergency department data may be valuable complement to analysis of inpatient mortality – in particular, do hospitals with high inpatient mortality also have relatively high rates of outpatient (emergency department) management, suggesting they are admitting a more severely ill subset of the stroke patients they treat?<br>• Chart reviews may be helpful to determine whether apparent mortality differences are the result of differences in coding, comorbidities, and/or disease severity. In particular, analysis should consider differences in hospital practices for using different ICD-9 codes related to |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | | | stroke.<br>• Validity of stroke results can also be examined by determining the sensitivity of estimated mortality rates to alternative plausible ICD-9 code definitions of new stroke patients (e.g., excluding patients undergoing endarterectomy on the admission date and excluding stroke-related codes with relatively low sensitivity and specificity for true strokes).<br>• Hospitals with many admissions for carotid endarterectomy and carotid angiography may have more "stroke" patients that do not represent true new strokes and that have better prognoses. |
| **In-Hospital Mortality after Surgical Procedures** | *Abdominal Aortic Aneurysm (AAA) Repair (#39)*<br><br>Lower AAA repair mortality may represent better quality. | • *Face validity:* AAA repair is a difficult procedure that carries a relatively high mortality rate. Evidence of a volume-outcome relationship and of better outcomes for technically proficient surgeons suggest that high-quality care can reduce AAA repair mortality.<br>• *Precision:* AAA repair is a relatively infrequent procedure, with a mean rate of well under 20 procedures per hospital per year. As a result, hospital mortality estimates that do not take account of random noise are likely to be unreliable.<br>• *Minimum bias:* Comorbid diseases measured in hospital discharge data as well as additional clinical factors (e.g., anesthesia risk and laboratory values) significantly influence postoperative | • *Precision*: Indicator demonstrates substantial variation in performance across providers. Most of the observed variation reflects unsystematic variation. Multivariate smoothing is particularly beneficial to this indicator.<br>• *Minimum bias*: Some of the observed variation in performance is due to systematic differences in patient characteristics. Provider rankings are somewhat affected by age-sex risk adjustment, suggesting that patient characteristics influence performance on indicator. Absolute performance is substantially impacted.<br>• *Construct validity*: AAA repair does not load substantially on any of the factors. It is correlated with other procedural mortality, such as CABG | HCUP Data Use:<br>• Though AAA mortality differs greatly among hospitals, most of the observed variation is random noise; "smoothed" estimates help avoid erroneous identification of "outlier" hospitals.<br>• Risk adjustment for APR-DRGs reduces random variations and bias, but has a relatively modest effect on hospital performance.<br><br>Further Investigation:<br>• If possible, useful to monitor outcomes for patients with AAA who do not undergo elective surgery, to assure that use of this QI does not lead to "access problems" and worse outcomes for difficult cases.<br>• It may be helpful to track mortality |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | mortality. No evidence on whether additional clinical measures influence hospital ranking.<br>• *Construct validity:* Volume, surgeon characteristics and practices, and associated clinical measures related to quality of care (e.g., surgical blood loss) have all been shown to be associated with significantly lower postoperative mortality at the patient level. Construct validity at the provider level not established.<br>• *Fosters true quality improvement*: Because unruptured AAA repair is an elective surgical procedure, its use as a quality measure could make it more difficult for high-risk patients to receive surgery. Use as a quality measure might also encourage earlier discharges or transfers.<br>• *Prior use:* This indicator is used by HealthGrades.com. Pennsylvania Health Care Cost Containment Council includes AAA mortality in a composite indicator for DRG 110. | and craniotomy mortality. | after hospitalization, though it is unclear that this is particularly beneficial for hip fracture, to avoid biases in apparent hospital mortality rates due to differences in hospital discharge and transfer practices.<br>• Medical chart reviews and analysis of laboratory data may be helpful both for determining whether more detailed clinical risk adjustment affects measured hospital performance and for identifying "best practices" associated with substantially lower mortality rates at some hospitals. |
| **In-Hospital Mortality after Surgical Procedures** | *Coronary artery bypass graft (CABG) (#40)*<br><br>Lower CABG mortality may represent better quality. | • *Face validity:* CABG is a complex procedure, and many technical processes of care as well as surgeon and hospital skill and experience have been shown to reduce mortality.<br>• *Precision:* CABG is a relatively common surgical procedure. However, a few studies have shown that random variations may cause a substantial number of hospitals to be mislabeled as outliers if "smoothing" methods are not used, because average mortality rates are not very high for most | • *Precision:* Indicator demonstrates adequate variation in performance across providers. Some of the observed variation reflects true differences in mortality, though some reflects unsystematic variation. Multivariate smoothing is particularly beneficial to this indicator.<br>• *Minimum bias:* Provider rankings are substantially affected by APR-DRG risk adjustment, suggesting that patient characteristics influence performance on indicator. In | HCUP Data Use:<br>• Although the provider-level mortality differences are somewhat smaller than for many other surgical procedures, the fact that CABG is very common makes this an important potential area for reducing mortality.<br>• Though significant differences in CABG mortality can be detected without use of statistical methods to account for random variation, some hospitals will be identified as |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | hospitals and because some hospitals perform relatively few procedures.<br>• *Minimum bias:* Patient demographics and comorbidities reported in discharge data, as well as additional clinical characteristics related to disease severity, are significant predictors of postoperative mortality. These factors do affect hospital rankings, although some evidence suggests that measured performance is relatively insensitive to which method is used to perform risk adjustment. Distinguishing comorbidities present on admission from complications of care can be difficult with discharge data.<br>• *Construct validity:* Evidence shows effect of volume, surgeon experience, on CABG mortality, though this evidence is equivocal. Aortic crossclamp or perfusion time has been shown to be associated with postoperative mortality, independent of disease severity.<br>• *Fosters true quality improvement:* Elective surgical procedure, so use as quality indicator may influence patients treated. Some evidence that use of CABG mortality as a QI has reduced surgical mortality, but it also may have reduced access to surgery for a very small number of patients.<br>• *Prior use:* One of the most commonly used indicators, CABG mortality is used by University Hospital Consortium, IMSystem, HealthGrades.com, Maryland QI | particular risk adjustment impacts providers with very high or low performance.<br>• *Construct validity:* Hospitals with higher CABG mortality also tend to perform worse on bilateratal catheterization and laparoscopic cholecystectomy. CABG mortality is significantly positively correlated with craniotomy mortality. | "outliers" as a result of random chance. Use of "smoothed" estimates will help avoid such errors.<br>• Risk adjustment, for APR-DRGs or additional clinical factors, appears to be quite important for this indicator.<br>• Consider in conjunction with other QI measures related to surgical procedures and cardiac care.<br><br>Further Investigation:<br>• It may be helpful to track mortality after hospitalization, to avoid biases in apparent hospital mortality rates due to differences in hospital discharge and transfer practices, though this has not been shown to be a particular problem for this indicator. Consider hospital average length of stay in interpreting HCUP measures.<br>• Medical chart reviews may be helpful both for determining whether more detailed clinical risk adjustment affects measured hospital performance; also might help distinguish complications from comorbidities.<br>• If possible, useful to monitor outcomes for patients with serious heart disease who do not undergo elective surgery, to assure that use of this QI does not lead to "access problems" and worse outcomes for difficult cases. |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | Project, Greater New York Hospital Association, Californ ia CABG Mortality Reporting Project, Cooperative Cardiovascular Project, Pennsylvania Health Care Cost Containment Council, New Jersey Department of Health and Senior Services, and New York Department of Health. | | |
| **In-Hospital Mortality after Surgical Procedures** | *Craniotomy mortality (#41)*<br><br>Lower craniotomy mortality may represent better quality. | • *Face validity:* Craniotomy for treatment of subarachnoid hemorrhage and cerebral aneurysm is a technically difficult procedure with significant postoperative risks of complications and mortality from strokes, infections, emboli, and other causes.<br>• *Precision:* This procedure is relatively common, with a hospital average rate of around 70 cases per year. Postoperative mortality is also relatively high, ranging from around 3% for young patients to around 15% for the elderly. Larger samples and higher event rates support more precise estimates.<br>• *Minimum bias:* Mortality risk varies substantially with patient age, which is correlated with (but may not completely account for) disease severity and likelihood of postoperative recovery. No evidence exists on the extent to which more detailed risk adjustment affects relative measures of provider performance.<br>• *Construct validity:* High procedure risk and evidence of a volume-outcome relationship suggest that | • *Precision:* Indicator demonstrates substantial variation in performance across providers. After smoothing, some of the observed variation reflects true differences in mortality, though some reflects unsystematic variation. Multivariate smoothing is particularly beneficial to this indicator.<br>• *Minimum bias:* Some of the observed variation in performance is due to systematic differences in patient characteristics. Provider rankings are somewhat affected by APR-DRG risk adjustment. In particular risk adjustment impacts providers with very high or low performance. Risk adjustment does change the absolute performance of providers substantially.<br>• *Construct validity:* Although craniotomy does not load on any of the factors, hospital mortality for craniotomy is positively correlated with AAA repair and CABG mortality. | HCUP Data Use:<br>• Large systematic differences across providers suggest that improving quality of craniotomy care may significantly improve outcomes.<br>• Use of "smoothed" estimates will help avoid errors in labeling hospitals as outliers, due to substantial random noise in apparent mo rtality rate differences across providers.<br>• Risk adjustment, including both demographics and APR-DRGs, is relatively important for this indicator.<br><br>Further Investigation:<br>• Medical chart reviews and laboratory tests may be helpful both for determining whether more detailed clinical risk adjustment affects measured hospital performance; also might help distinguish complications from comorbidities.<br>• If possible, useful to monitor outcomes for patients with subarachnoid hemorrhage or cerebral aneurysm who do not undergo elective surgery, to assure that use of this QI does not lead to "access problems" and worse outcomes for difficult |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | hospitals can reduce mortality by directing procedures to experienced surgeons and by improving processes of care. However, construct validity based on other process or outcome measures has not been established at either the patient or provider level.<br>• *Fosters true quality improvement:* Elective surgical procedure, so use as quality indicator may influence patients treated and their discharge management.<br>• *Prior use:* This indicator is used by University Hospital Consortium. | | cases. |
| **In-Hospital Mortality after Surgical Procedures** | *Esophageal resection mortality (#42)*<br><br>Lower esophageal resection mortality may represent better quality. | • *Face validity*: Very difficult procedure in patients with typically complex underlying disease, and high postoperative mortality, suggesting that provider skill and use of appropriate processes of care may significantly reduce mortality.<br>• *Precision*: Very infrequent procedure, with most hospitals performing less than five procedures per year. Even though mortality is common and appears to differ substantially based on provider volume, small numbers suggest that random variations at the provider level may be very important.<br>• *Minimum bias*: Although risk adjustment for demographics and comorbid diseases may help explain variation in patient outcomes, no evidence on whether better risk adjustment alters measured provider performance significantly.<br>• *Construct validity*: Providers with higher volumes have lower mortality | • *Precision*: Indicator demonstrates adequate variation in performance across providers. Most of the observed variation reflects unsystematic variation. Multivariate smoothing is beneficial to this indicator.<br>• *Minimum bias*: Some of the observed variation in performance is due to systematic differences in patient characteristics. Provider rankings and absolute performance are both somewhat affected by age-sex risk adjustment.<br>• *Construct validity*: Due to the distribution of this indicator, construct validity was not evaluated. | *HCUP Data Use:*<br>• Large systematic differences across providers suggest that improving quality of care for esophagectomy patients may significantly improve outcomes, though relatively few patients would be affected.<br>• Use of "smoothed" estimates will help avoid errors in labeling hospitals as outliers, due to substantial random noise in apparent mortality rate differences across providers.<br>• Risk adjustment, including both demographics and APR-DRGs measured in discharge data, does not appear to alter provider performance measures substantially, except for providers with extreme low and high rates.<br>• Consider in conjunction with other QI measures related to complex cancer surgery procedures (e.g., pancreatic resection). |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | rates.<br>• *Fosters true quality improvement*: Because virtually all hospitals perform only a small number of esophageal resections, changing practices for a relatively small number of patients (e.g., avoiding a few cases each year with the highest mortality risk) may significantly alter a hospital's performance measure.<br>• *Prior use*: This indicator is discussed in the literature as a volume-outcome measure. | | *Further Investigation:*<br>• It may be helpful to track mortality after hospitalization, to avoid biases in apparent hospital mortality rates due to differences in hospital discharge and transfer practices.<br>• Medical chart reviews and laboratory tests may be helpful both for determining whether more detailed clinical risk adjustment affects measured hospital performance, and for identifying processes of care and patient characteristics that lead to much lower mortality rates. |
| **In-Hospital Mortality after Surgical Procedures** | *Hip replacement mortality (#43)*<br><br>Lower acute hip replacement mortality may represent better quality. | • *Face validity*: Hip replacement is a common procedure frequently performed on older patients with comorbid illnesses, and many technical processes of care may reduce potentially life-threatening complications.<br>• *Precision*: Hip replacement is a relatively common surgical procedure, with hospitals performing over 70 per year on average. However, in-hospital mortality rates are low, often much less than 1-2%. As a result, hospital-level estimates may be imprecise for many hospitals.<br>• *Minimum bias*: Patient demographics and serious comorbidities (particularly the occurrence of hip fracture) are significant predictors of postoperative mortality. Limited evidence suggests that including additional clinical risk adjusters affects rankings for some hospitals.<br>• *Construct validity*: Some evidence | • *Precision*: Indicator demonstrates adequate variation in performance across providers. Most of the observed variation reflect unsystematic variation, not actual differences in mortality. Multivariate smoothing is not particularly beneficial for this indicator.<br>• *Minimum bias*: Some of the observed variation in performance is due to systematic differences in patient characteristics. Provider rankings and absolute performance are both substantially affected by APR-DRG risk adjustment, suggesting that patient characteristics influence performance on indicator. In particular risk adjustment impacts providers with very high or low performance.<br>• *Construct validity*: Hip replacement is not substantially correlated with any of the other adult procedures, except craniotomy. | *HCUP Data Use:*<br>• Although hospital mortality rates for this procedure are generally low, some hospitals have substantial mortality that is not explained by random variations, suggesting that important reductions in hip fracture mortality may be possible.<br>• Because most of the variation in hip fracture mortality across hospitals can be attributed to random chance, some hospitals will be identified as "outliers" unless methods such as "smoothing" are used to account for such errors. However, multivariate smoothing does not improve the signal for this indicator substantially.<br>• Risk adjustment with APR-DRG at minimum is recommended for this indicator, due to evidence of potential bias.<br>• Consider in conjunction with other QI measures related to surgical procedures and medical conditions |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | suggests that better thromboembolic prophylaxis leads to lower mortality, though evidence on the association with other processes of care and whether such processes can explain hospital-level outcome differences is very limited. Some studies show an association between hospital volume of hip replacements and outcomes, possibly indicating that better processes of care or experience lead to better outcomes.<br>• *Fosters true quality improvement*: Elective surgical procedure, so use as quality indicator may influence patients treated and the timing of their hospital discharge. | | that are common in the elderly.<br><br>*Further Investigation:*<br>• Medical chart reviews may be helpful both for determining whether more detailed clinical risk adjustment affects measured hospital performance.<br>• Use as a quality indicator may discourage surgeons from taking difficult cases. To assess this possibility, clinical data could be used to determine whether operative risk declines in conjunction with the adoption of the measure. |
| **In-Hospital Mortality after Surgical Procedures** | *Pancreatic resection mortality (#44)*<br><br>Lower pancreatic resection mortality may represent better quality. | • Essentially same as esophageal resection | • *Precision*: Indicator demonstrates substantial variation in performance across providers. Most of the observed variation reflects unsystematic variation. Multivariate smoothing is particularly beneficial to this indicator.<br>• *Minimum bias*: Some of the observed variation in performance is due to systematic differences in patient characteristics. Provider rankings and absolute performance are both substantially affected by age-sex risk adjustment, suggesting that patient characteristics influence performance on indicator. In particular risk adjustment impacts providers with very high or low performance.<br>• *Construct validity*: Due to the distribution of this indicator, construct validity was not evaluated. | • Essentially same as esophageal resection. |
| **In-Hospital** | *Pediatric heart* | • *Face validity*: Pediatric heart surgery | • *Precision*: Indicator demonstrates | *HCUP Data Use:* |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| **Mortality after Surgical Procedures** | *surgery mortality (#45)*<br><br>Lower pediatric heart surgery mortality may represent better quality. | involves complex operations that require considerable technical skill.<br>• *Precision*: Pediatric heart surgery is already concentrated in a limited number of specialized facilities, though measures may be imprecise for some centers that have relatively low volumes.<br>• *Minimum bias*: Pediatric heart surgery consists of a wide range of procedures associated with very different operative risks. Mortality for particular procedures also varies widely according to patient clinical characteristics (baseline functional status and comorbidities, cardiac anatomy, etc.). Such factors could confound hospital mortality comparisons, though little evidence exists on this question.<br>• *Construct validity*: Some evidence suggests that better technique (e.g., greater technical skill and experience leading to shorter cardiopulmonary bypass time) significantly reduces mortality. While no studies have documented an association between hospital mortality measures and such processes of care, some studies have found a volume-outcome relationship that may reflect better processes and experience.<br>• *Fosters true quality improvement*: Some procedures are elective, so use as quality indicator may influence patients treated. However, the indications for most types of pediatric heart surgery are relatively clear, so | adequate variation in performance across providers. Most of the observed variation reflect unsystematic variation, not actual differences in mortality. Multivariate smoothing is particularly beneficial to this indicator.<br>• *Minimum bias*: Some of the observed variation in performance is due to systematic differences in patient characteristics. Provider rankings are substantially affected by APR-DRG risk adjustment, suggesting that patient characteristics influence performance on indicator.<br>• *Construct validity*: Though this indicator does not load on any of the factors, pediatric heart surgery is negatively correlated with pediatric heart surgery volume. | • Hospital variations in mortality are substantial, even after risk adjustment based on discharge data, though most of the apparent variation can be attributed to random noise.<br>• Because most of the variation in pediatric heart surgery mortality across hospitals can be attributed to random chance, some hospitals will be identified as "outliers" unless methods such as "smoothing" are used to account for such errors.<br>• Some risk adjustment is possible using hospital discharge data. However, many important risk adjusters are not included in discharge data, and there is little evidence on whether accounting for such factors would affect measured hospital performance.<br>• Because this indicator consists of a broad range of surgical procedures, additional investigation of unusual mortality rates might focus on differences in surgical mix and/or on mortality for particular types of surgical procedures. However, even at large centers, the number of cases for such procedures is small enough that random noise is likely to be a major problem.<br>• Consider in conjunction with other mortality measures for other major surgical procedures and serious medical conditions.<br><br>*Further Investigation:*<br>• Medical chart reviews may be helpful |

| Indicator Type | Indicator | Literature review findings | Empirical performance | Suggestions for use |
|---|---|---|---|---|
| | | that more difficult patients might end up being referred to lower-mortality hospitals. Use of this measure may also affect the timing of hospital discharges, with unclear implications for outcomes. | | for determining whether more detailed clinical risk adjustment affects measured hospital performance.<br>• If possible, helpful to track mortality after hospitalization, to avoid biases in apparent hospital mortality rates due to differences in hospital discharge and transfer practices. Consider hospital average length of stay and transfer rates in interpreting HCUP measures.<br>• Use as a quality indicator may discourage surgeons from taking difficult cases. To assess this possibility, clinical data could be used to determine whether operative risk declines in conjunction with the adoption of the measure, whether more difficult cases were increasingly referred to lower-mortality centers, and whether outcomes for children with serious heart defects improved. |

## 3.A.4. Summary Tables of Evidence by Empirical Tests

### Volume Indicator Results

Statistical tests for volume indicators differ from utilization, admission and mortality rates, as precision of the indicator is not a primary concern. The following section summarizes the results of each statistic derived for the volume indicators. Some of the statistics were not used in the "rating" of an indicator, and are simply provided for context; this is noted where applicable. This section is organized into three tables, each summarizing one of the three areas of empirical performance for volume indicators, distribution, share, and persistence (see section 2.E. "Empirical Methods" for a full explanation of these methods). A total of 906 hospitals were included in this evaluation. Two thresholds were used for each indicator, with the exception of pediatric heart surgery, where only one threshold has been recommended. The thresholds were derived using thresholds reported in the literature and represent less and more stringent criteria for high volume.

**Table 14. Thresholds for indicators**

| | Thresh. 1 | Literature source(s) | Thresh. 2 | Literature source(s) |
|---|---|---|---|---|
| Carotid endarterectomy | 50 | Manheim, 1998[188] | 101 | Hannan, 1998; Dudley, 2000[189, 190] |
| PTCA | 200 | Ryan, 1993[191] | 400 | Hannan, 1997; Kimmel, 1995[61, 192] |
| CABG | 100 | Eagle, 1995[193] | 200 | Hannan, 1991[54] |
| Pediatric heart surgery | 100 | Hannan, 1998[194] | - | - |
| AAA repair | 10 | Hannan, 1992[195] | 32 | Kazmers, 1996; Provonost, 1999[196, 197] |
| Esophageal resection | 6 | Patti, 1998[198] | 7 | Dudley, 2000; Patti, 1998[190, 198] |
| Pancreatic resection | 10 | Glasgow, 1996[199] | 11 | Glasgow, 1996[199] |

**Table 15. Volume distribution by year**

| | Volume distribution in 1996 | | | | | |
|---|---|---|---|---|---|---|
| | % Hosp with ≥1 procedure / year | Mean | SD | 50th | 90th | 95th |
| Carotid endarterectomy | HIGH | 52 | 60 | 31.5 | 129 | 169 |
| PTCA | LOW | 418[a] | 400 | 330 | 869 | 1157 |
| CABG | LOW | 399[a] | 338 | 293 | 830 | 1095 |
| Pediatric heart surgery | MODERATE | 53 | 90 | 2.5 | 149 | 245 |
| AAA repair | HIGH | 14 | 16 | 8 | 36 | 46 |
| Esophageal resection | LOW | 2 | 3 | 1 | 4 | 6 |
| Pancreatic resection | MODERATE | 3 | 4 | 2 | 5 | 10 |

[a]Volume for PTCA is increasing, with a mean of 379 procedures in 1995. Volume for CABG is increasing with a mean of 375 in 1995.

Table 15 outlines the volume distribution for each procedure (listed in the left hand column). It includes 6 statistics:

***Percent of hospitals performing one or more procedures annually.*** No procedure was performed by all hospitals. Since many hospitals do not perform the procedure in question, this indicator does not apply to

those specified hospitals. This statistic gives the percentage of hospitals that perform this procedure and thus are impacted by this indicator. If only a small number of providers perform a procedure, the face validity of that indicator, as it pertains to the importance of that procedure, is called into question. However, other aspects of face validity and construct validity are important as well, and were weighted more strongly when evaluating the indicator.

HIGH: More than 50%
MODERATE: 25% - 50%
LOW: Less than 25%

***Mean number of procedures performed annually.*** This statistic is the simple mean of procedures performed annually by hospitals that perform at least one procedure. If the mean number of procedures is low, the precision for that indicator may be affected. This statistic was not weighted, but rather is reported for context.

Tests 3-6 provide additional information on the distribution of procedures across hospitals. Examining the distribution highlights cases where very few providers perform a majority of procedures, or other special cases that would result in a skewed distribution.
***Standard deviation (SD).*** Standard deviation of the number of procedures performed annually. This statistic includes only hospitals performing the procedure.
***50th percentile.*** Median of the number of procedures performed annually. This statistic includes only hospitals performing the procedure.
***90th percentile*** *of the number of procedures performed annually*. This statistic includes only hospitals performing the procedure.
***95th percentile*** *of the number of procedures performed annually*. This statistic includes only hospitals performing the procedure.

**Table 16. Percentage of procedures and volume**

| Percentage of procedures at high volume hospitals in 1996: | | |
|---|---|---|
| | Threshold 1 | Threshold 2 |
| Carotid endarterectomy | MODERATE | MODERATE |
| PTCA | HIGH | MODERATE |
| CABG | HIGH | HIGH |
| Pediatric heart surgery | MODERATE | - |
| AAA repair | MODERATE | LOW |
| Esophageal resection | LOW | LOW |
| Pancreatic resection | LOW | LOW |
| **Percentage of hospitals at high volume in 1996:** | | |
| | Threshold 1 | Threshold 2 |
| Carotid endarterectomy | 36.6% | 16.7% |
| PTCA | 68.9% | 42.1% |
| CABG | 87.6% | 67.8% |
| Pediatric heart surgery | 20.6% | - |
| AAA repair | 44.4% | 12.2% |
| Esophageal resection | 8.6% | 6.4% |
| Pancreatic resection | 5.1% | 4.2% |

Table 16 summarizes the percentage of procedures and volume in two statistics. Both statistics are reported for two thresholds. Threshold 1 is the less stringent threshold, and threshold 2 is the more stringent threshold. Thresholds are reported at the beginning of this section, in table 14.

***Percentage of procedures at high volume hospitals.*** This statistic refers to the number of procedures performed at hospitals that qualify as high volume (according to the thresholds) as a percentage of all procedures. If most procedures are already performed at high volume centers, there is little room for improvement (few procedures could be shifted to high volume hospitals). On the other hand, if most procedures are performed at low volume hospitals, many procedures could be shifted to high volume hospitals, potentially resulting in better outcomes. Therefore, a designation of “low” implies the most opportunity for improvement.

Threshold 1:
HIGH: More than 90%
MODERATE: 50%-90%
LOW: Less than 50%

***Percentage of hospitals at high volume*.** This statistic refers to the number of hospitals performing more than the threshold number of procedures as a percentage of all hospitals performing at least one procedure. This statistic was not weighted in the evaluation, but rather is provided for context in interpreting the first statistic.

**Table 17. Year to year persistence of high volume status**

| | Percentage of Hospitals at High Volume Remaining at High Volume (1996-1997) | |
|---|---|---|
| | Threshold 1 | Threshold 2 |
| Carotid endarterectomy | MODERATE[a] | MODERATE |
| PTCA | HIGH | HIGH |
| CABG | HIGH | HIGH |
| Pediatric heart surgery | MODERATE | - |
| AAA repair | MODERATE | MODERATE |
| Esophageal resection | LOW | LOW |
| Pancreatic resection | LOW | MODERATE[a] |

[a] For 1995 to 1996, CE has high persistence (threshold 1). During the same years, pancreatic resection has low persistence (threshold 2).

Table 17 demonstrates the persistence of the high volume status of hospitals from year to year. Reported is the percentage of hospitals designated as high volume (according to the thresholds) in 1995 or 1996, that remain high volume the following year. If high volume status is not persistent, it is likely that simply examining one year of data may not be accurate. Indicators with low persistence are likely to have lower precision.

HIGH: More than 90%
MODERATE: 75% - 90%
LOW: Less than 75%

**Table 18. Correlations between volume and mortality indicators [a]**

| | Vol-CE | Vol-PTCA | Vol-CABG | Vol-Pedi heart surg | Vol-AAA repair | Vol-Esoph resect | Vol-Panc resec | Mort-Esoph resect | Mort-Pancr resect | M A r |
|---|---|---|---|---|---|---|---|---|---|---|
| Volume-Carotid Endarterectomy N = 1429 | 1 | | | | | | | | | |
| Volume-PTCA N = 1429 | .59 | 1 | | | | | | | | |
| Volume-CABG N = 1429 | .61 | .78 | 1 | | | | | | | |
| Volume-Pediatric heart surgery N = 1429 | .12 | .12 | .18 | 1 | | | | | | |
| Volume-Abdominal aortic aneurysm repair N=1429 | .61 | .46 | .45 | .19 | 1 | | | | | |
| Volume-Esophageal resection N = 1429 | NS | NS | .13 | .18 | NS | 1 | | | | |
| Volume-Pancreatic resection N=1429 | .14 | .12 | .15 | .24 | .18 | .40 | 1 | | | |
| Mort-Esophageal resection N = 42 | NS | NS | NS | NS | NS | **-.29** | -.34[b] | 1 | | |
| Mort-Pancreatic resection N = 118 | NS | NS | NS | NS | NS | -.19[b] | **-.41** | .48[b] | 1 | |
| Mort-Abdominal aortic aneurysm repair N = 534 | -.38 | -.45 | -.43 | NS | **-0.35** | NS | -.13[b] | NS | NS | |
| Mort-Pediatric heart surgery N=57 | NS | NS | NS | **-.27[b]** | NS | NS | NS | NS | NS | |
| Mort-CABG N = 286 | -.26 | -.21 | **-0.29** | NS | -.18[c] | NS | NS | NS | NS | |

[a]Unless otherwise noted, correlations are significant at p<.001. N.S. indicates a non-significant correlation. Correlations of volume and mortality for the same procedure are bolded.
[b]significant at the p <.05 level
[c]significant at the p = .01 level

Table 18 displays the Pearson r correlations between volume and mortality indicators significant at the p<.001. Where previous hypotheses suggest a potential relationship, significant correlation at the p<.05 level are reported. As the validity of volume indicators is based on the volume-outcome relationship, we tested the correlation between volume and mortality. As we did not use the same methods (i.e. clinical risk adjustment) as most of the studies cited in the literature reviews, results from this test are not considered conclusive. Rather, the literature review informs the construct validity for these indicators.

## Precision Results

### Table 19. Precision – Provider Indicators [a,b]

| Indicator[c] | Mean | Std Dev | Std Dev (1) | Share (2) | Ratio (3) | R-Sq (4) |
|---|---|---|---|---|---|---|
| CESAREAN SECTION DELIVERY (#8)[d] | 0.214 | 0.087 | HIGH | HIGH | HIGH | VERY HIGH |
| INCIDENTAL APPENDECTOMY (#9) | 0.027 | 0.035 | MODERATE | HIGH | MODERATE | MODERATE |
| Bi-lateral Catheterization (#10) | 0.193 | 0.200 | VERY HIGH | VERY HIGH | VERY HIGH | VERY HIGH |
| VAGINAL BIRTH AFTER C-SECTION (#11)[d] | 0.336 | 0.148 | VERY HIGH | HIGH | HIGH | HIGH |
| LAPAROSCOPIC CHOLECYSTECTOMY (#12)[d] | 0.662 | 0.192 | VERY HIGH | VERY HIGH | HIGH | HIGH |
| Mort-AMI (#33) | 0.244 | 0.161 | HIGH | MODERATE | MODERATE | MODERATE |
| Mort-Congestive Heart Failure (#34) | 0.075 | 0.095 | MODERATE | MODERATE | MODERATE | MODERATE |
| Mort-GI Hemorrhage (#35) | 0.046 | 0.057 | MODERATE | MODERATE | LOW | MODERATE |
| Mort-Hip Fracture (#36) | 0.144 | 0.160 | HIGH | VERY HIGH | MODERATE | MODERATE |
| Mort-Pneumonia (#37) | 0.138 | 0.102 | HIGH | HIGH | MODERATE | HIGH |
| Mort-Stroke (#38) | 0.213 | 0.137 | HIGH | HIGH | MODERATE | HIGH |
| Mort-Abdominal aortic aneurysm repair (#39)[d] | 0.215 | 0.268 | HIGH | HIGH | LOW | MODERATE |
| Mort-CABG (#40) | 0.051 | 0.062 | MODERATE | MODERATE | MODERATE | MODERATE |
| Mort-Craniotomy (#41) | 0.162 | 0.185 | HIGH | HIGH | LOW | MODERATE |
| Mort-Esophageal resection (#42)[d] | 0.202 | 0.366 | MODERATE | MODERATE | LOW | LOW |
| Mort-Hip Replacement (#43) | 0.012 | 0.057 | MODERATE | HIGH | LOW | LOW |
| Mort-Pancreatic resection (#44)[d] | 0.154 | 0.313 | HIGH | HIGH | LOW | LOW |
| Mort-Pediatric heart surgery (#45) | 0.072 | 0.170 | MODERATE | MODERATE | LOW | LOW |

[a]See "key" below for explanation of table.
[b]All results, except for raw mean and standard deviation, reflect APR-DRG risk adjusted data, except where otherwise noted.
[c]Indicators in all capital letters are part of HCUP I QI set.
[d]Age-sex risk adjusted only.

**Table 20. Precision – Area Indicators [a,b]**

| Indicator[c] | Mean | Std Dev | Std Dev (1) | Share (2) | Ratio (3) | R-Sq (4) |
|---|---|---|---|---|---|---|
| CORONARY ARTERY BYPASS GRAFT (#13) | 0.001804 | 0.005716 | VERY HIGH | VERY HIGH | VERY HIGH | VERY HIGH |
| HYSTERECTOMY (#14) | 0.004194 | 0.003233 | VERY HIGH | HIGH | VERY HIGH | VERY HIGH |
| LAMINECTOMY AND/OR SPINAL FUSION (#15) | 0.001390 | 0.003475 | HIGH | VERY HIGH | VERY HIGH | VERY HIGH |
| PTCA (#16) | 0.001908 | 0.004556 | VERY HIGH | VERY HIGH | VERY HIGH | VERY HIGH |
| Dehydration (#17) | 0.001399 | 0.001032 | MODERATE | MODERATE | HIGH | HIGH |
| Bacterial Pneumonia (#18) | 0.003956 | 0.002085 | HIGH | HIGH | VERY HIGH | VERY HIGH |
| Urinary Infection (#19) | 0.001451 | 0.000895 | MODERATE | MODERATE | HIGH | HIGH |
| PERFORATED APPENDIX (#20) | 0.332740 | 0.144097 | VERY HIGH | VERY HIGH | LOW | LOW |
| Angina (#21) | 0.001660 | 0.001357 | HIGH | HIGH | VERY HIGH | VERY HIGH |
| Adult Asthma (#22) | 0.001079 | 0.000817 | MODERATE | MODERATE | HIGH | HIGH |
| COPD (#23) | 0.003240 | 0.002038 | HIGH | HIGH | VERY HIGH | VERY HIGH |
| Congestive Heart Failure (#24) | 0.005210 | 0.002865 | HIGH | HIGH | VERY HIGH | VERY HIGH |
| DIABETES SHORT TERM COMPLICATION (#25) | 0.000360 | 0.000246 | MODERATE | MODERATE | MODERATE | MODERATE |
| Diabetes Uncontrolled (#26) | 0.000347 | 0.000281 | MODERATE | MODERATE | HIGH | HIGH |
| DIABETES LONG TERM COMPLICATION (#27) | 0.000808 | 0.000581 | MODERATE | MODERATE | HIGH | HIGH |
| Hypertension (#28) | 0.000371 | 0.000322 | MODERATE | MODERATE | MODERATE | HIGH |
| Lower Extremity Amputation (#29) | 0.000305 | 0.000427 | MODERATE | MODERATE | MODERATE | HIGH |
| LOW BIRTH WEIGHT (#30) | 0.039108 | 0.023135 | VERY HIGH | VERY HIGH | MODERATE | HIGH |
| PEDIATRIC ASTHMA (#31) | 0.001541 | 0.001439 | HIGH | HIGH | HIGH | HIGH |
| Pediatric Gastroenteritis (#32) | 0.000985 | 0.001012 | HIGH | MODERATE | HIGH | HIGH |

[a]See "key" below for explanation of table.
[b]All results, except for raw mean and standard deviation, reflect age-sex risk adjusted data.
[c]Indicators in all capital letters are part of HCUP I QI set.

**Key to Empirical Evaluation of QI Precision**

(1) Standard Deviation:
The standard deviation of the "signal" in each QI is a measure of the extent to which performance on the QI varies systematically across hospitals or areas. It is a best estimate of the variation in QI performance that appears to be systematic – truly associated with the hospital or area, and not the result of random variations in patients or environmental conditions. Because many patient characteristics and random events may influence performance on a QI, only a fraction of the apparent variation in performance at the hospital or area level is likely to be systematic. Systematic variation will be larger to the extent that: sample sizes are larger (allowing more precise estimation of the true effect); patient-associated variation is smaller (allowing more precise estimates for a given sample size); and other factors discussed in more detail in the text. Larger variation across hospitals and areas in the signal suggests that there is truly more variation across hospitals or areas to explain and, possibly, to improve. In contrast, if the standard deviation is small, the QI's performance suggests there is little to be gained by improving the performance of lower-ranked hospitals or areas to the levels achieved by the higher-ranked hospitals or areas. The standard deviations in the table are reported as absolute rates. For example, for VBAC, the mean rate across all providers and years is 32.3% - just under one-third of patients reported to have a previous C-section underwent vaginal delivery. The systematic variation across hospitals in VBAC performance is large – the standard deviation of the estimated performance of hospitals around the overall average is 17.5%. In other words, the precision of the VBAC QI is sufficient to conclude that many hospitals truly had VBAC rates close to 0, whereas many others truly had VBAC rates over 50%. QIs for hospitals and areas were grouped into three broad performance categories based on the absolute magnitude of the true or signal standard deviation (in percentage points):

Provider QI:
VERY HIGH – 8.0%+
HIGH – 3.0% to 7.9%
MODERATE – Less than 3.0%

Area QI:
VERY HIGH – 0.15%+
HIGH – 0.05% to 0.15%
MODERATE – Less than 0.05%

(2) Share:
The share reported in the Tables is the share of signal variance (that is, the standard deviation squared) in the *total* variance associated with the QI. In general, the variance in a quality indicator can be divided as follows:
Total Variance = Patient-Level Variance + Provider-Level Variance.
Typically, much variation in the numerators for the QIs (e.g., whether or not a vaginal birth occurred for a patient with a prior C-section) is associated with patient-level factors, and appears to have nothing whatsoever to do with providers or areas. In turn, the apparent Provider-Level Variance can be divided into two components:
Provider-Level Variance = Signal Variance + Noise Variance.
That is, some of the variance in a QI rate for a particular provider (adjusted or unadjusted) is a reflection of random chance: in a given year, some hospitals or areas will by chance have more or fewer

occurrences of the numerator related to a particular QI. The remaining variance, and only the remaining variance, is attributable to systematic or true differences in the QI across providers. The share of signal variance will be larger, to the extent that: within-hospital patient variance is small; sample sizes are larger; and true differences in performance across hospitals are larger. Unlike the Standard Deviation reported in the previous column – see (1) above – the share of signal variance is a relative measure: relative to the total variation (patient plus provider) in the QI, what fraction appears to be associated with systematic provider differences? Other things equal, higher shares suggest differences across hospitals or areas that will be easier to sort out from all of the other "random" influences on QI performance. QIs for hospitals and areas were grouped into three broad performance categories in terms of the signal share of total variation:

Provider QI:
VERY HIGH – 6.0%+
HIGH – 1.0% to 5.9%
MODERATE – Less than 1.0%

Area QI:
VERY HIGH – 0.10%+
HIGH – 0.03% to 0.10%
MODERATE – Less than 0.03%

(3) Ratio:
The ratio reports the Signal-to-Noise ratio in the *provider*- or *area*- level variation in the QI measure (see Note 2 for a discussion of the sources of variation in a performance measure). This measure answers the question: of the apparent variation in QIs across providers, what fraction appears to be *truly* related to systematic differences across providers, and not random variations ("noise") from year to year? As such, the measure is somewhat misnamed. Its definition is:
Signal-to-Noise Ratio = (Provider-Level Signal Variation)/(Provider-Level Variation)
where (as noted above) the provider-level variation includes both signal variation and random variations in the QI measures across providers ("noise"). In other words, it is the "signal" to "signal plus noise" ratio. In general, if a QI's signal-to-noise ratio is high, then it is likely that apparent variations in performance across providers are not the result of random chance, and careful attention to distinguishing true from random variation across providers will have little impact on the measured performance of a provider or area. QIs for hospitals and areas were grouped into three broad performance categories in terms of the signal share of provider-level variation:

VERY HIGH – 90.0%
HIGH – 70.0% to 89.9%
MODERATE – 40.0% to 70.0%
LOW – Less than 40%

(4) R-square:
The above ratio (3) reports the univariate smoothed ratio. However, multivariate techniques may improve the ability to recover signal. An R-square that is higher than the ratio demonstrates the ability of multivariate techniques to extract more signal than univariate techniques. The same rating criterion used for the ratio is used for the R-square measure.

Section 3.A.4. Summary Results by Empirical Tests

How can these measures be used to guide QI initiatives?

1. Signal Standard Deviation: If there is little true variation across providers or areas, then there is little to be gained by comparing the performance of providers or areas as a basis for understanding differences in performance and possibly improving poor performance. Without further refinement, computing the QI is unlikely to have much policy value. Many potential QIs show very little true variation, and so were dropped from further analysis. Virtually all of the indicators included in our recommended list had a standard deviation of at least 1 percentage point in performance, that is, the top 2.5% of hospitals or areas performed at least 4 percentage points above the lowest 2.5% of hospitals or areas.

2. Share of Signal in Total Variation: If the share of total variation is very small, it is likely to be more difficult as a practical matter to identify why hospitals or areas perform well or poorly. In other words, because so little of the variation in the QI at the patient level is explained, there will probably be many potential explanations for differences in performance that may be quite costly to pursue. As a cautionary note, it is important to recognize that the bulk (over 90%) of the variation in the numerators for virtually all of the QIs we reviewed is not systematic provider-level variance. Thus, identifying the reason(s) why a particular hospital or area performed well or poorly on the QI is generally likely to be difficult. Our review of the literature on the QIs provides some further guidance for promising specific explanations to pursue for particular QIs. If determining the cause of systematic differences appears to be difficult, the QI may not be worth pursuing.

3. Signal to Noise Ratio in Provider Variance: Even if a QI performs reasonably well in terms of signal standard deviation and share of total variation, it may still have a low share of true provider-level differences in the apparent differences across providers. In this case, careful statistical analysis is required before a provider or area that appears to perform well or poorly can be labeled as such. As our empirical evaluation describes, this further analysis might consist of evaluating measures over multiple years, evaluating measures that are related to each other, and applying "Bayesian" techniques to account for the fact that the apparent differences across providers are "noisy." In this table, we report the signal variation that all such methods can recover; the signal variation in a single, univariate analysis is likely to be smaller.

Section 3.A.4. Summary Results by Empirical Tests

## Minimum Bias Results

**Table 21. Minimum Bias – Provider Indicators [a,b]**

| Indicator[c] | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| CESAREAN SECTION DELIVERY[d] (#8) | VERY GOOD | VERY GOOD | GOOD | FAIR | VERY GOOD |
| INCIDENTAL APPENDECTOMY (#9) | VERY GOOD | VERY GOOD | GOOD | GOOD | VERY GOOD |
| Bi-lateral Catheterization (#10) | VERY GOOD | VERY GOOD | GOOD | VERY GOOD | VERY GOOD |
| VAGINAL BIRTH AFTER C-SECTION[d] (#11) | VERY GOOD | VERY GOOD | VERY GOOD | GOOD | VERY GOOD |
| LAPAROSCOPIC CHOLECYSTECTOMY[d] (#12) | VERY GOOD | VERY GOOD | VERY GOOD | FAIR | VERY GOOD |
| Mort-AMI (#33) | FAIR | FAIR | FAIR | FAIR | FAIR |
| Mort-Congestive Heart Failure (#34) | GOOD | GOOD | FAIR | FAIR | FAIR |
| Mort-GI Hemorrhage (#35) | GOOD | GOOD | FAIR | FAIR | FAIR |
| Mort-Hip Fracture (#36) | GOOD | FAIR | GOOD | FAIR | FAIR |
| Mort-Pneumonia (#37) | FAIR | FAIR | FAIR | FAIR | FAIR |
| Mort-Stroke (#38) | GOOD | GOOD | GOOD | FAIR | FAIR |
| Mort-Abdominal aortic aneurysm repair (#39)[d] | GOOD | FAIR | GOOD | FAIR | GOOD |
| Mort-CABG (#40) | FAIR | GOOD | FAIR | FAIR | FAIR |
| Mort-Craniotomy (#41) | GOOD | FAIR | FAIR | FAIR | FAIR |
| Mort-Esophageal resection (#42)[d] | GOOD | GOOD | GOOD | VERY GOOD | GOOD |
| Mort-Hip Replacement (#43) | FAIR | FAIR | FAIR | FAIR | FAIR |
| Mort-Pancreatic resection (#44)[d] | FAIR | FAIR | GOOD | FAIR | FAIR |
| Mort-Pediatric heart surgery (#45) | FAIR | GOOD | FAIR | FAIR | FAIR |

[a]See "key" below for explanation of table.
[b]All indicators are APR-DRG risk adjusted, unless otherwise noted.
[c]Indicators in all capital letters are part of HCUP I QI set.
[d]Demographic (age-sex) risk adjustment only.

**Table 22. Minimum Bias – Area Indicators [a,b]**

| Indicator[c] | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| CORONARY ARTERY BYPASS GRAFT (#13) | FAIR | GOOD | FAIR | VERY GOOD | FAIR |
| HYSTERECTOMY (#14) | VERY GOOD | VERY GOOD | GOOD | GOOD | VERY GOOD |
| LAMINECTOMY AND/OR SPINAL FUSION (#15) | GOOD | VERY GOOD | FAIR | VERY GOOD | GOOD |
| PTCA (#16) | FAIR | GOOD | FAIR | VERY GOOD | FAIR |
| Dehydration (#17) | VERY GOOD | VERY GOOD | GOOD | GOOD | VERY GOOD |
| Bacterial Pneumonia (#18) | GOOD | GOOD | GOOD | GOOD | GOOD |
| Urinary Infection (#19) | GOOD | GOOD | GOOD | GOOD | GOOD |
| PERFORATED APPENDIX (#20) | VERY GOOD | VERY GOOD | VERY GOOD | GOOD | VERY GOOD |
| Angina (#21) | VERY GOOD | GOOD | GOOD | GOOD | VERY GOOD |
| Adult Asthma (#22) | VERY GOOD | VERY GOOD | VERY GOOD | VERY GOOD | VERY GOOD |
| COPD (#23) | GOOD | GOOD | GOOD | GOOD | GOOD |
| Congestive Heart Failure (#24) | GOOD | GOOD | FAIR | FAIR | FAIR |
| DIABETES SHORT TERM COMPLICATION (#25) | VERY GOOD | VERY GOOD | VERY GOOD | VERY GOOD | VERY GOOD |
| Diabetes Uncontrolled (#26) | VERY GOOD | VERY GOOD | GOOD | GOOD | VERY GOOD |
| DIABETES LONG TERM COMPLICATION (#27) | GOOD | GOOD | GOOD | GOOD | GOOD |
| Hypertension (#28) | VERY GOOD | VERY GOOD | GOOD | VERY GOOD | VERY GOOD |
| Lower Extremity Amputation (#29) | GOOD | GOOD | GOOD | GOOD | GOOD |
| LOW BIRTH WEIGHT[d] (#30) | NA | NA | NA | NA | NA |
| PEDIATRIC ASTHMA (#31) | VERY GOOD | VERY GOOD | VERY GOOD | VERY GOOD | VERY GOOD |
| Pediatric Gastroenteritis (#32) | VERY GOOD | VERY GOOD | VERY GOOD | VERY GOOD | VERY GOOD |

[a]See "key" below for explanation of table.
[b]All area indicators are adjusted by age and sex only
[c]Indicators in all capital letters are part of HCUP I QI set.
[d]Adequate risk adjustment unavailable.

**Key to Empirical Evaluation of QI Bias:**

To provide empirical evidence on the sensitivity of our QIs to potential biases 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 severity. We used the APR-DRG System (3M Version 12.0, severity and risk-of-mortality subclasses) for adjustment (See literature review for discussion). For a few measures, no APR-DRG categories were available, so that unadjusted measures were compared to age-sex adjusted measures. 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, it is possible and in some cases likely that there is not 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 specific indicators. The table reports results for our multivariate signal extraction models, which generally yielded the most precise estimates of QI performance measures for areas and hospitals. Results for other statistical methods were largely similar. For all five performance measures, described below, we classified performance into three groups.

(1) Rank Correlation Coefficient: This is the correlation coefficient of the rank of the area/hospital without and with risk adjustment.

VERY GOOD – 95.0%+
GOOD – 75.0% to 94.9%
FAIR – Less than 75.0%

(2) Average Absolute Value of Change Relative to Mean: This is a measure of the average absolute change (in percentage points) in the performance measure of the hospital or area with and without risk adjustment, normalized by the average value of the QI. Thus, it is also a relative measure that theoretically can range from 0 (no change) to a much higher value.

VERY GOOD – Less than 10.0%
GOOD – 10.0% to 20.0%
FAIR – 20.0%+

(3, 4) Percentage of High/Low Decile That Remains in High/Low Decile: These two measures report the percentage of hospitals or areas that are in the highest and lowest performance deciles without risk adjustment, which remain there with risk adjustment. Thus, a measure that is insensitive to risk adjustment should have rates of 100%, while measures where risk adjustment affects the top- and bottom-performers substantially will have much lower rates. (Note: “Top 10%” for all measures refers not to the best performers, but to the hospitals or areas with the *top (highest)* rates. For some measures, such as VBAC, this may indicate better performance; for others, such as the mortality measures, it generally indicates worse performance.) Given the distributions of the indicators, it is much “easier” to move out of the highest decile relative to the lowest decile. For this reason different rating criteria are used.

Section 3.A.4. Summary Results by Empirical Tests

HIGH DECILE:
VERY GOOD – 85.0%+
GOOD – 55.0% to 84.9%
FAIR – Less than 55.0%

LOW DECILE:
VERY GOOD – 95.0%+
GOOD – 85.0% to 94.9%
FAIR – Less than 85.0%

(5) Percentage That Change More Than Two Deciles: This measure reports the percentage of hospitals whose relative rank changes by a substantial distance – more than 20% - with and without risk adjustment. Ideally, risk adjustment would not have a substantial effect for very many hospitals.

VERY GOOD – Less than 5.0%
GOOD – 5.0% to 14.9%
FAIR – 15.0%+

## Construct Validity Results

While most evidence on the construct validity of our indicators is based on our review of the literature, our empirical analysis allows us to determine the extent to which hospital and area performance measures are related to each other. Consistency of hospital performance across measures can provide some empirical evidence on construct validity: one might expect that hospitals with better underlying processes of care will tend to perform better on a range of quality measures. It is not necessarily the case that hospitals performing better in some aspects of care will perform better on others, and there is evidence that good performance in some aspects of care is not related to good performance in others. However, many of these prior studies used performance measures that included substantial measurement error, and possibly other biases; other things equal, such "noise" will cause relationships among measures to appear attenuated. Thus, these studies may have underestimated the extent of relationships across measures and the potential for achieving more general insights about quality from the joint analysis of specific quality indicators.

To quantify the extent to which performance on specific quality indicators reflected general patterns, we conducted factor analyses using as inputs our best-guess estimates (those obtained through the MSX analysis) of hospital and area performance on the full set of quality indicators that had adequate precision. Though esophageal and pancreatic resection mortality both had adequate precision, very few hospitals actually perform these procedures regularly, resulting in a non-normal distribution. Thus, these indicators were excluded, as they violated the normality assumption of factor analysis.

Factor analysis seeks to summarize results for a large number of specific measures using a smaller number of summary indicators, or factors. It does so by obtaining estimates of the "loadings" of each specific quality measure on each summary factor, in order to explain the maximum variation in all of the quality measures possible with a limited number of factors. Quality measures closely related to each other will have high "loadings" on a particular factor. Quality measures with significantly positive loadings on a particular factor tend to occur together; conversely, a quality measure with a significant negative loading tends to be inversely related to quality measures that have positive loadings on a factor. To minimize attenuation bias caused by imprecision in the indicators, we conducted this analysis using the "best estimates" of the hospital quality indicators obtained through the multivariate signal extraction (MSX) method.

We report the results of three separate common factor analyses. The three factor analyses included 1) *hospital* analysis with all hospital-level quality indicators; 2) *area* factor analysis with all of the area-level indicators related to the quality of ambulatory care; 3) *area* analysis with all of the area-level utilization indicators. Because of the large number of hospitals and areas, all with an inherently large sample size, all factor loadings were significant, even if the associations are not very substantial. We therefore examined the factors with relation to strength of factor loadings, directions of loadings, and the explanatory value of a factor for any given quality indicator. For each quality indicator, we identified the factor that explained the largest proportion of the variance. In addition to the factor analyses we examined correlation (Pearson's r) matrixes for each group (provider level, ACSC, and utilization area indicators) to aid us in interpreting the factor analysis.

To highlight the most important findings from our factor analysis, we identified the most important explanatory factor for each quality indicator, that is, the factor that explained the largest share of the variation across areas or providers in the indicator. We report both factor loadings (using a varimax orthogonal rotation), which correspond to the correlation between that indicator and that factor, and standardized factor loadings, calculated by multiplying the factor loading by the standard deviation for that indicator and multiplying by 100. This results in a measure of percent change, meaning a 1 unit

change in the factor would result in a 1 percentage point change in the indicator. We used a cutoff of a standardized loading of 0.7, which corresponds to an explanation of about half of the variance in the quality indicator, as our measure of "important." Tables 23-25 summarize the results.

***Provider-level analysis.*** Indicators related to both mortality and utilization were included in our hospital analysis. We identified three factors, using a criterion of Eigenvalue greater than 1 to extract factors. Many indicators loaded on more than one factor with substantial strength, confirming that the determination of hospital performance on the quality measures is a complex process that is not easily summarized by any one factor. Nonetheless, the three factors together provide a relatively complete summary of hospital performance.

Factor 1 explained most of the variance for most of the medical mortality measures (stroke, pneumonia, hip fracture, CHF, GI hemorrhage), although this factor explains less than 50% of the variance for the latter. As expected these indicators are all positively related to each other, meaning hospitals with high mortality rates for one condition, tend to have high mortality rates for the other conditions. Factor 1 explains 39% of the total common variance.

Factor 2 explained 19% of the total common variance. It explains most of the variance for two cardiac care indicators, mortality after CABG, and bilateral catheterization, as well as laparoscopic cholecystectomy. As expected the two cardiac care indicators are positively related to each other, while negatively related to laparoscopic cholecystectomy.

The final factor, factor 3, explains most of the variance for the obstetric utilization indicators, VBAC and C-section rates, and for the other non-technology based utilization factor, incidental appendectomy. As might be expected, VBAC and C-section have a significant negative relationship with each other: hospitals with high C-section rates tend to have low VBAC rates. Incidental appendectomy is positively related to cesarean section and negatively related to VBAC, as one would expect. Factor 6 explains 13% of the total common variance.

While the strongest relationships in each of the factors seem to follow a logical pattern based on quality relationships, there are also less important associations that do not follow this pattern. For instance, better performance in some quality measures on these factors was negatively correlated with other quality measures. There are many reasons for such imperfect relationships, such as negative correlations conditional on the other factors (e.g., the third factor should be considered in light of the first two and so forth), unexplained associations despite our best efforts to remove noise from the measures, and validity problems in the quality indicators (e.g., failure of the indicators to describe true quality differences accurately, due to possible residual biases as well as limited construct validity). However, overall, the most important relationships shown in the factor analyses are readily interpretable using clinical logic.

Several factors did not load strongly on any of the three factors, including all of the procedural mortality measures. These indicators did tend to be positively related to each other, as observed from the correlation matrix and factors. However, in general, the correlations were less strong than those for the medical mortality measures.

Given our time and resource constraints, we were unable to perform a full set of evaluations of our factor analysis results. However, these results suggest some important conclusions. Despite quite noisy data, most medical mortality indicators were positively related to each other, suggesting that some underlying "mortality factor" explains an important component of the variation in hospital performance on the quality measures.

Second, technology based utilization factors of laparoscopic cholecystectomy and bilateral catheterization were negatively related to each other, and CABG mortality is positively related to

bilateral catheterization, suggesting a possible underlying "best practices" factor. Finally, the C-section and VBAC rates at a hospital are strongly negatively related; this would not be expected if variation in VBAC rates across hospitals primarily reflected reporting differences rather than true differences in how procedures were used. Incidental appendectomy is also related as would be expected to VBAC and C-section, if there were an underlying high quality "factor" as it relates to non-technology based practices.

**Table 23. Summary of Factor Analysis Results, Provider-level indicators [a]**

| | Name of indicator | Direction of loading | Strength of factor loadings[b] | Standardized strength of factor loadings[c] | % of total common variance explained by factor |
|---|---|---|---|---|---|
| **Factor 1** | Mortality - Stroke | + | high | very high | 38.9% |
| | Mortality - Pneumonia | + | high | very high | |
| | Mortality - GI Hemorrhage | + | very high | moderate | |
| | Mortality - Congestive Heart Failure | + | high | high | |
| | Mort – Hip Fracture | + | moderate | high | |
| **Factor 2** | Bi-lateral Catheterization | + | high | very high | 19.4% |
| | Mortality - CABG | + | high | high | |
| | Laparoscopic Cholecystectomy | - | high | very high | |
| **Factor 3** | Vaginal Birth after C-section | + | very high | very high | 25.2% |
| | Cesarean section | - | high | very high | |
| | Incidental Appendectomy | - | high | high | |

[a]Indicators shown represent those for which the given factor explains most of the within -indicator variation.
[b]Varimax rotation raw factor loadings. Very high: |.76 - .99|; high: |.51-.75|; moderate: |.35 – .50|.
[c]Standardized factor loadings. A factor loading of 1 represents a 1 percent point change in indicator with a 1 unit change in the factor. Very high: higher than |2|; high: |0.7| - |1.99|; moderate: |0.5| - |0.7|

***Ambulatory care sensitive condition (ACSC) analysis.*** This analysis of the relationship among the area-level ACSC measures showed that, for the most part, rates of these conditions are positively correlated with each other. Two factors were extracted from this analysis, using the criterion Eigenvalue greater than 1.

Factor one explains a large proportion of the common variance, over 70%. It also explains the most variance for each of the factors except one, lower extremity amputation. All, but two indicators, low birth weight, and perforated appendectomy loads positively on this indicator. Both of these indicators are defined with a specified denominator (percentage of births, appendectomies, respectively), and appear to be different from the other ACSC indicators. These two indicators also load lower than the other indicators on this factor. Pediatric asthma also loads only slightly on this factor, using the raw varimax rotated factor loadings.

Factor two only explains 22% of the common variance. Lower extremity amputation loads the highest on this factor, with a factor loading of .90. Factor two also explains the most variance for this indicator. For the most part all of the other indicators load much weaker on factor two, especially compared to their loadings on factor one. Five exceptions are pediatric asthma, diabetes long term complications, diabetes short term complications, CHF, and low birth weight. These indicators are all positively related to each other. However, factor 1 explains more of the variance than does factor two for these indicator. Subsequent factors should be interpreted in light of previous factors, especially given the large difference in variance explained by the factors in this case.

Overall, this factor analysis shows that most ACSC indicators, particularly those included in nationally used sets (Billings and Weissman), are significantly and positively related to each other, as construct validity would suggest.

Section 3.A.4. Summary Results by Empirical Tests

**Table 24. Summary of Factor Analysis Results, Area-level ACSC indicators [a]**

| | Name of indicator | Direction of loading | Strength of factor loadings[b] | Standardized strength of factor loadings[c] | % of total common variance explained by factor |
|---|---|---|---|---|---|
| **Factor 1** | Bacterial pneumonia | + | Very high | Very high | 70.7% |
| | Congestive heart failure | + | High | Very high | |
| | COPD (chron. obst. pulm. disease) | + | Very high | Very high | |
| | Angina | + | High | Very high | |
| | Dehydration | + | High | High | |
| | Urinary infection | + | High | High | |
| | Adult asthma | + | High | High | |
| | Pediatric gastroenteritis | + | Moderate | High | |
| | Pediatric asthma | + | Low | Moderate | |
| | Uncontrolled diabetes | + | Very high | Moderate | |
| | Hypertension | + | Very high | Moderate | |
| | Diabetes long term complications | + | Moderate | Moderate | |
| | Lower extremity amputation[d] | + | - | - | |
| | Diabetes short term complications | + | Moderate | Low | |
| | Perforated appendix | - | Low | Very high | |
| | Low birth weight | - | Low | Very high | |

[a]Indicators shown represent those for which the given factor explains most of the within -indicator variation.
[b]Varimax rotation raw factor loadings. Very high: |.76 - .99|; high: |.51-.75|; moderate: |.25 – .50|.
[c]Standardized factor loadings. A factor loading of 1 represents a 1 percent point change in indicator with a 1 unit change in the factor. Very high: higher than |.051|; high: |.025| - |.050|; moderate: |.010| - |.024|
[d]Factor two explains more of the variance for lower extremity amputation than factor one.

***Area utilization indicator analysis.*** Again, one factor was sufficient to explain a large share of the total variation in our area utilization indicators (over 96%). All utilization indicators load on one factor in the same direction, as expected.

Overall, the factor analysis of area-level utilization rates shows that the rates are generally positively related to each other, suggesting that any of these indicators provides insights into the other utilization rates.

**Table 25. Summary of Factor Analysis Results, Area-level utilization indicators**

| | Name of indicator | Direction of loading | Strength of factor loadings[a] | Standardized strength of factor loadings[b] | % of total common variance explained by factor |
|---|---|---|---|---|---|
| **Factor 1** | CABG | + | Very high | Very high | 96.7% |
| | Laminectomy and/or spinal fusion | + | Very high | High | |
| | PTCA | + | Very high | High | |
| | Hysterectomy | + | High | Moderate | |

[a]Varimax rotation raw factor loadings. Very high: |.76 - .99|; high: |.51-.75|; moderate: |.25 – .50|.
[b]Standardized factor loadings. A factor loading of 1 represents a 1 percent point change in indicator with a 1 unit change in the factor. Very high: higher than |.50|; high: |.25| - |.50|; moderate: |.10| - |.24|

## 3.B. Overview of Semi-Structured Interviews

The semi-structured interviews were conducted to understand the implications on the HCUP QI refinement project of current quality measurement practices in the health care field. Those interviewed cited the use of several types of indicators as an alternative or in addition to the HCUP measures. Many organizations are using length of stay and cost proxies as indicators. Readmission rates, ambulatory care sensitive conditions and patient satisfaction surveys were also mentioned repeatedly. Several organizations utilize proprietary measures, such as the Maryland Quality Indicator Project. Most organizations utilized risk adjusted measures, often using 3M APR-DRGs as a risk adjustment method.

The HCUP I QIs are known by most of the organizations. They appear to be best known by state data groups and hospital associations, and least known by business coalitions. Several organizations are currently using the HCUP I QIs for a variety of purposes, including benchmarking, reference for decision-making, and community health reporting.

There were many suggestions for modifications and additions to the HCUP I QIs, detailed in Table 26. We briefly summarize those suggestions and issues raised that were most relevant to refinement of the current HCUP I QIs. Areas of concern regarding the current HCUP I QIs include the following:

- Indicator denominators: The indicator denominators sometimes consist of heterogeneous populations. Some organizations would like to see further modification of the denominators, while others wish to see population based denominators instead of admission based denominators.
- Low volume: Many of the events examined by the HCUP I QIs are rare events, such as the mortality measures. This is a particular problem for hospitals with low patient volume to begin with such as rural hospitals. Suggestions for this concern included creating optional aggregate or basket measures to increase volume, and including diagnoses with higher mortality rates such as hip pinnings or unruptured aneurysms.
- Validity: Some organizations expressed concern over the validity of the assumptions in the indicator definitions. For instance, some argue that one cannot assume that coded complications were not present on admission. Some organizations suggested that this concern be addressed through sixth digit coding or other additional data.
- Data integrity: There is some concern over coding bias in the discharge abstract. Some state data groups raised the issue of data audits as a way to ensure data integrity.

One of the most common suggestions for refinement of the HCUP I QIs was risk adjustment. Organizations seem to have accepted risk adjustment as a standard in quality measurement. Many indicated that they preferred risk adjustment, as opposed to the narrowing of populations to create a more homogenous comparison group. Many organizations currently use the 3M APR-DRGs for risk adjustment. These organizations have indicated their preference to continue using risk adjustment.

In addition to refining the current HCUP I measures, organizations also indicated areas of potential novel indicators. For the most part many of these mirrored the expansion areas defined by AHRQ, including chronic disease and new technologies. Other clinical domains included adverse selection, mental health, and volume sensitive procedures.

Some suggested new indicators for the HCUP II QI set that would require the use of additional data, including at least the possibility of linking data with records such as census records, vital statistics and outpatient records. Several states are currently collecting additional data on ambulatory surgery, such that

it may be possible to define indicators in this domain. The use of these data, while beyond the scope of the current project, would increase the flexibility of the HCUP II QIs.

Those interviewed emphasized the need for the HCUP II QIs to remain flexible. There was great interest in the use of HCUP II QIs in conjunction with other indicators. Organizations expressed that the most useful new developments would interact with the measures they currently use, allowing the organizations to create a more complete picture of quality. Contacts' suggestions included designing HCUP II to interact with patient satisfaction measures that are currently widely used, such as the FAACT surveys, or the Picker survey.

Further pointing to the need for flexibility is the wide range of audiences with possible interests in the HCUP II QIs. As mentioned earlier our contact list included a variety of organizations with a variety of goals. Some organizations, such as state data groups and hospital associations, want to benchmark hospitals, while others want to report on community health. Other organizations, such as business coalitions, seek to make purchasing decisions. These differing goals led to some differing opinions regarding indicators, such as the ambulatory care sensitive conditions. For instance, one organization valued the ambulatory care sensitive conditions for county-wide reporting of health. Other interviewees expressed that these indicators are not valid indicators of hospital quality since hospitals may not be responsible for these events. Thus, the range of uses and consequent opinions need to be considered when refining the HCUP QIs.

**Table 26. Summary of Telephone Contacts August 1999 through March 2001**

| Response | Detailed Comment |
|---|---|
| **Current Indicator Use** | |
| CAHPS (Consumer Assessment of Health Plans) | AHCPR developed survey instrument;<br>FACCT developed complement for children's ambulatory and hospital care (Medicaid and Children's Health Insurance Program (CHIP)) |
| Picker Institute | The Picker surveys target: access to care, respect for patients' values, coordination of care, information and education, involvement of family and friends, physical comfort, emotional support, and transition and continuity of care.<br>The inpatient survey covers: Adult medical/surgical, Childbirth, Comprehensive maternity, Inpatient rehabilitation, Pediatric inpatient, Neo-natal intensive care unit (NICU) |
| Maryland QI Project | Overlapping with current HCUP I measures, proprietary system<br>QI set includes: Peri-operative mortality, cesarean section rate, VBAC rate<br>Web-based interface for data entry and analysis |
| Hospital Reporting of QIs | Public (New Mexico)<br>Not Public (New York)<br>Statewide only (Rhode Island) |
| **Current HCUP I Indicator Evaluations and Suggestions** | |
| Changes to Indicator Denominators | Obstetrical complications (surgical patients only); Wound infection (surgical patients only); Adverse effects and iatrogenic complications (surgical patients only); Utilization indicators (sensitive to adverse selection; population) |
| Inclusion of Population Denominators | Community level<br>Number of conditions per 100,000 population (how to define market share of hospitals remains a difficulty) |
| Low Volume of events used in indicators | Events for following indicators too rare for meaningful hospital comparisons, especially among rural hospitals:<br>In-hospital mortality following common elective procedures (hip/knee); Low birth weight; Very low birth weight; Laminectomy mortality; cholecystectomy mortality. |
| Summary/ aggregate measures | Problem of integration, which indicators should be combined?; Low volume of events used in indicators drives the need for summary measures. Aggregate measures are preferable to pooling years of data on a single indicator. Quality "CPI" |
| Validity | Concern about validity of complications among surgical patients. |
| HCUP I Complications among Surgical Patients | Second most widely used after mortality, including: AMI after major surgery, GI hemorrhage, pulmonary compromise, pneumonia.<br>Other measures include length-of-stay, costs as proxy (with or without mortality).<br>"Present on admission" data element would be helpful.<br>Independence of "effectiveness" and "efficiency" affects the validity of these indicators.<br>False assumption in HCUP I logic that surgical patients would not under go treatment in presence of complication.<br>Need data of time of occurrence of complication for profiling. |
| HCUP I Coded Complications | Heterogeneous denominator; Should examine relation to mortality.<br>Other measures include length-of-stay, costs as proxy (with or without mortality).<br>These indicator hold the least confidence among HCUP I QIs due to 5th digit reporting and other coding practices.<br>Independence of "effectiveness" and "efficiency" affects the validity of these indicators. |

Section 3.B. Results of Semi-structured Interviews

| Response | Detailed Comment |
|---|---|
| HCUP I Mortality | Most widely used indicators. Often severity-adjusted with 3M APR-DRG system. Actual rates are very low for current indicators. Others possible indicators include non-ruptured aneurysms, CABG, hip pinning. |
| HCUP I Utilization | Least widely used indicators. Some concerns about the validity of the CABG, Cholecystectomy measures.<br>"-ectomy" measures of political interest (unnecessary procedures).<br>Single hospital rates subject to hospital selection; population rates preferable. Subject to patient preferences. Severity-adjustment problematic. |
| Ambulatory Care Sensitive Conditions | Would be better if more hospitals had target markets. Difficult for hospitals to accept accountability; medical groups may be more accountable. Suggest county level reporting (diabetes, asthma). Alternative to Codman/Wennberg.<br>Useful for the identification of access to care problems (e.g., state-funded community clinics).<br>Provider level discharge denominator subject to hospital selection. |
| State data audits | Would prefer audit standards (similar to NCQA). Need to identify data elements to audit. |
| **Severity/Risk Adjustment** | |
| Severity Adjustment – current use | Most widely used: 3M APR-DRGs<br>Others risk adjustment systems noted: Medstat (Disease Staging), Atlas (score 1-5 based on chart review within 48 hours of admission; presence of cancer). |
| Severity Adjustment – preferences for implementation | Prefer severity adjustment to removing complicated cases to achieve homogeneity.<br>Mortality and measures of intensity: charges, readmissions, length of stay, medical DRGs, surgical DRGs<br>Prefer system based on common terminology, developed with non-elderly and elderly, open system, performs well on mortality.<br>Important for communication to hospitals, physicians, employers.<br>Recommendations for whether certain HCUP I QIs lend themselves to certain severity adjustment systems would be helpful. |
| **Addition of Novel Indicators and Data Elements** | |
| Clinical Domains | Additional clinical domains to examine:<br>Nutrition and nutritional deficiencies (elderly); Access (Acute ambulatory care); Chronic conditions; Immunizable conditions; Short-stay hospitalizations; Short-stay surgeries (and complications) |
| New Technologies | Ideas for incorporating measures of new technology:<br>Retrospective index of 'technology leader' based on the early adoption of new technologies subsequently widely adopted<br>"Economies of scale" based on high-volume, specialized practices |
| Adverse Selection | Identified as a potential indicator. |
| Mental Health/ Rehabilitation | Measures of delay in treatment; truncated LOS<br>Large share of hospitals licensed by states |
| Volume sensitive procedures | Potential candidates (CABG, NICU, trauma) |
| Linked data | At minimum it would be helpful to include logic for linked admission records for states with data<br>Can use linked data for the validation of discharge-based indicators (maternity, time-to-readmit)<br>Encourage states to adopt linked data |
| Vital Statistics | At minimum it would be helpful to include logic for linked death records for states with data<br>Can be used for testing of alternative time-to-death (discharge, surgery)<br>Provides location of death, cause of death |

Section 3.B. Results of Semi-structured Interviews

| Response | Detailed Comment |
|---|---|
| | Encourage states to adopt |
| Data elements | 6th digit coding for complication and co-morbidity can be used for testing "condition present on admission"<br>Testing 'plan type'<br>Supporting other UB-92 enhancements |
| Ambulatory Surgery | Additional measures. Data collection efforts in Wisconsin, New York, Pennsylvania, Connecticut, New Jersey |
| Flexibility of Indicator Set | |
| Dartmouth Atlas | Integrate Atlas with HCUP II QIs to improve visibility of HCUP II indicator set. |
| Sixth Scope of Work (1999-2002) | HCFA Peer Review Organizations will collect data on performance measures in six clinical area: acute myocardial infarction, congestive heart failure, pneumonia, stroke/atrial fibrillation, diabetes, and breast cancer.<br>Emphasis on comparisons: state-to-state, Fee for Service versus managed care, urban versus rural |
| Benchmarking | State-to-state comparisons (age-sex standardized)<br>Hospital peer groups<br>Benchmark data available from vendors (HCIA). Concern over the timeliness of NIS benchmarks. |
| Purchasers | Increase HCUP II visibility<br>Public purchaser groups (e.g., Calpers)<br>Quality metrics, such as QIs, useful for value-based purchasing<br>Complement HEDIS, FACCT, CAHPS |
| Joint Commission | Important that HCUP II measures are acceptable as ORYX indicators<br>Alternative to Maryland QI<br>Economic consequence (accreditation) is most important for adoption of indicator sets. |
| AHRQ Imprimatur | Acceptance by hospitals and state legislators |

# 3.C. Review of Risk Adjustment Approach

This section presents the results of a literature review on potential risk adjustment systems and presents the evidence leading to the selection of APR-DRGs. All risk adjustment systems have potential drawbacks, and much has been written regarding risk adjustment in general. In the first section, this evidence is discussed to delineate the general caveats regarding risk adjustment. Further, any risk adjustment system used in conjunction with the HCUP QIs could only use the data available in the HCUP data set. This eliminated many risk adjustment systems that require additional clinical data. In the second section, we discuss the comparative evidence on systems that apply to ICD-9-CM coded administrative data. The final section discusses the potential of risk adjustment systems to fulfill the user defined criteria outlined in section 2.D. "Risk adjustment of HCUP Quality Indicators."

## Literature Review of Risk Adjustment Systems

***General Caveats About Risk-Adjustment Systems***. For a risk factor to affect inter-hospital or -area comparisons of performance on a quality indicator, two conditions must be met. First, the likelihood that the patient will experience the outcome of interest must differ based on the presence or absence of the risk factor. Second, the distribution of the risk factor must differ systematically across the relevant patient populations treated by particular hospitals (if this condition does not hold, then better risk adjustment improves the precision of the quality indicator but will not change relative performance).[200] Analysis of the extent to which each of these conditions is true should ideally take place after accounting for random variations across hospitals or areas; otherwise, random variations in patient mix may obscure systematic differences across particular hospitals or areas. If *both* of these conditions hold, then differences across hospitals or areas in performance on an indicator will reflect true differences in provider quality *plus* these systematic patient differences. Thus, the systematic differences in patient disease severity lead to *bias* in the performance measure. Risk adjustment systems seek to account for measurable differences in patient risk factors in comparisons across areas and providers, to remove bias caused by these patient differences.

Several general comments apply to all the risk adjustment systems that might be used to adjust quality indicators. First, all available risk adjustment systems explain a relatively small share of the variation in virtually all hospital performance measures, whether measures of health outcomes and resource use. In other words, the proportion of outcomes at the individual level predicted correctly and the variation in outcomes "explained" by the risk factors included in the risk adjustment systems is much less than 100 percent. This is especially true for risk adjustment systems based on administrative data only, but is also true for systems based on much more detailed clinical data. As a result, risk adjustment systems generally do not help to improve the precision of comparisons across providers very much (by reducing the unexplained variance in outcomes across providers). If comparisons across providers are imprecise, the value of risk adjustment will tend to be lower, because it is difficult to detect significant differences in provider performance. In addition, while it is possible that the remaining unexplained variation is truly random, it is also possible that additional differences in patient characteristics across providers or areas are not captured by the risk adjustment system. Such residual differences in patient mix remain a source of bias even after risk adjustment. Because adjustment systems

based on more detailed clinical data generally have more discriminatory power, they will generally lead to less biased comparisons. Conversely, inaccurate or inconsistent recording of risk factors that are included in a risk adjustment system can increase noise and bias in comparisons across providers. Because such anomalies are more likely to arise in risk adjustment systems with complex and burdensome reporting requirements, relying on risk factors that are clearly and consistently measurable may lead to less biased comparisons in practice.

Do risk adjustment systems based on careful reviews of medical charts and other supplemental, costly data sources result in substantially different conclusions about performance than systems based on administrative data? For some measures, collecting additional clinical information on risk factors not measured in discharge data is clearly valuable: it substantially alters the relative and absolute conclusions about provider performance.[201, 202] For many other measures, however, the impact on comparisons across providers appears to be minimal.[203, 204] Again, this may be because the additional variation explained by the additional risk factors is small, or because the additional risk factors do not differ much across the providers or areas being compared (e.g., only a small share of patients may have the comorbidity or severity factors). Because this evidence is to a large extent measure-specific, we generally discuss it in the context of the reviews of evidence on particular quality indicators.

Four caveats to these conclusions should be noted. First, models that incorporate laboratory data, which are increasingly available through hospital information systems, may perform nearly as well as models based on more detailed clinical data, at substantially lower cost.[19] Indeed, the correlation of hospital-level predicted values between models based on administrative plus laboratory data, and models based on full clinical data, has been reported to be as high as 0.98.[205] Second, confusion between comorbidities and complications may compromise the validity of risk-adjustment based on administrative data, relative to risk-adjustment based on detailed clinical data.[19,112] An indicator variable designating conditions that were present at admission (currently available only in California and New York) may greatly enhance the validity of risk-adjustment by allowing analysts to avoid misadjusting for conditions that actually represent adverse outcomes.[113] Third, the relative superiority of models based on clinical data depends somewhat on which specific clinical data elements are available and how they are used. The number of clinical data elements that actually contribute to explaining outcomes at the provider level may be relatively small;[201 206] failing to collect these key confounders may lead one to the incorrect conclusion that clinical data are unnecessary.[207] Finally, because much variation in performance is unexplained by even the most detailed risk adjustment methods, it is possible that all available risk adjustment methods fail to account fully for important, systematic differences in risk across hospitals.

**Evidence regarding DRG based systems.** Several systems use the type of discharge data readily available in HCUP. These include PIP-DCG, Medstat's Disease Staging, and DRG based systems such as HCFA DRGs, AP-DRGs and APR-DRGs. Although a substantial literature exists for Medstat's Disease Staging and although DxCG's PIP-DCG is being implemented to adjust payments to private plans in Medicare, none of the other systems are in use so universally with state data organizations as DRG based systems. Such official recognition, and lower cost of implementing a system that is already in use proved important to surveyed users of the HCUP I QIs. Resource constraints prohibited us from evaluating HCUP II QIs under each of these systems individually, thus our evaluation focused mainly on the DRG based

systems. Future evaluations may consider other potential systems using the methods applied here.

Because other systems are also based on the DRG framework, it is worth considering what advantages the specific implementation of the APR-DRG system offers. We reviewed a recent comparison of DRG systems[208] in the context of potential HCUP II QIs. As mentioned earlier, there are two separate severity classifications in APR-DRG, one for severity of illness (resource use and complications) and one for risk of mortality. The other DRG systems considered were the Medicare DRGs, Refined DRGs (a HCFA-funded project at Yale University), All-Patient DRGs (jointly developed by the New York Department of Health and 3M), and Severity DRGs (another HCFA-sponsored refinement from the mid-1990s). The primary distinguishing characteristics of the APR-DRG system compared to these other DRG-based systems were: 1) the inclusion of severity categories that reflect not only presence of complications and co-morbidities, but also an assessment of the level of these conditions; 2) a severity classification for particular secondary diagnoses that varies by principal condition, age, and operative procedure; this approach recognizes a differential impact of secondary diagnoses by condition (i.e., primary diagnosis) and allows for some interactions among secondary diagnoses.

*Literature review results.* In empirical analyses (assumably as a result of these distinguishing characteristics discussed above), the APR-DRG system performed better than alternative DRG-based systems in explaining variation in both cost and mortality at the patient and provider levels. Specifically, in explaining patient level cost variation, the APR-DRG system produced a higher R-squared (explained more of the variance) than the alternative systems, primarily as a result of the additional explanatory power of the secondary diagnosis subgroups. The impact of the secondary diagnosis subgroups was greater for medical patients than surgical patients, especially for non-normal neonates, pediatric patients with chronic conditions, patients older than 65, and patients treated in children's hospitals and major teaching hospitals. The importance of hospital characteristics in explaining cost variation was less for the APR-DRG system than the other DRG systems, reflecting that the system captures the more complex case-mix at these providers.

Similarly, in explaining patient level mortality variation, the APR-DRG system produced a higher R-squared and c-statistic than the alternative systems, and resulted in better calibration for major diagnostic categories and age and sex subgroups. As in the cost analysis, the secondary diagnosis subgroups were primarily responsible for the increase. However, in contrast to the cost analysis, for mortality the impact of the secondary diagnosis subgroups was slightly greater for surgical patients than medical patients. The importance of hospital characteristics in explaining mortality variation was less for the APR-DRG system than the other systems, reflecting that the system captures the more complex case-mix at certain providers. The greatest impact of the DRG refinements occurred for large urban and teaching hospitals; the least impact occurred for small urban and rural hospitals. It is important to note that teaching hospitals may also have more severe case mixes or may systematically code more comorbidities than smaller hospitals.

These findings are largely consistent with our empirical analysis of the impact of the APR-DRG risk adjustment system on relative provider performance among the HCUP II QIs (see below). For example, the impact of APR-DRG risk adjustment was greater for surgical patients than for medical patients in provider level mortality measures, suggesting that the measures for procedures (which are largely elective) may be subject to more selection bias

related to inpatient mortality than most of the "medical" measures. The impact tended to be greater for hospitals with measured performance in the highest and lowest decile. This was consistent with the Averill's finding that "trimming" of outlier cases tended to mute the impact of the APR-DRG severity classification system relative to the other DRG-based systems.[208] In other words, compared to the alternative DRG systems, the severity levels in the APR-DRG system do a better job of reflecting the distribution of patient severity at the extremes, where accounting for case-mix differences may be most important for the application of QIs.

A series of studies by Iezzoni and colleagues have also compared the performance of APR-DRGs (in a prior version) and other severity-adjustment systems based on administrative data (See Table 27). They reported that APR-DRGs performed better than competing systems (Disease Staging mortality probability, Patient Management Categories severity scale) in predicting inpatient mortality after coronary artery bypass graft surgery and stroke, but worse than Disease Staging in predicting inpatient mortality after acute myocardial infarction and pneumonia. Even in comparison with Refined DRGs, APR-DRGs had superior discrimination for some conditions (e.g., acute myocardial infarction, stroke), but inferior discrimination for other conditions (e.g., pneumonia, coronary artery bypass graft surgery). [9-13, 17, 18, 204] These findings suggest that: (1) No severity-adjustment system, based on either administrative or clinical data, is clearly superior for all conditions and procedures. The optimal system for one condition may not be optimal for another. It is impossible to predict in advance, without empirical analysis, which system would be superior for any specific condition. (2) APR-DRGs perform about as well, and in some cases better, than competing severity-adjustment systems based on administrative data. (3) Unfortunately, different severity-adjustment systems frequently produce different impressions about severity-adjusted performance at the provider level, even when these systems have comparable discrimination.

In addition, DRGs are one of the few risk adjustment systems based on administrative data that have been used in conjunction with "smoothing" methods – the statistical methods intended to remove the effects of random variations on measured differences in performance across providers and areas. Burgess et al.[209] illustrate the use of DRGs in this way. As in our analysis, the investigators applied age, sex, and DRG-based risk adjustment to hospital performance measures along with statistical methods for smoothing out random variations in measures (empirical Bayes methods). They found that hospitals with relatively imprecise performance measures (in particular, hospitals with small numbers of patients) were more likely to perform closer to the overall average performance level once the methods were applied. In other words, smoothing helped reinforce the effects of severity-adjustment in removing differences in patient mix that led smaller providers to appear to be outliers, when in fact they were not. We apply similar methods in our evaluation of the HCUP II QIs to obtain similar "smoothed" estimates of risk-adjusted provider performance.

**Table 27. Risk adjustment statistics from the literature**

| | AMI (N=11,880) | | | CABG (N=7,765) | | |
|---|---|---|---|---|---|---|
| **Severity Measure** | **C** | **$R^2$** | **% deaths deciles 1-10** | **C** | **$R^2$** | **% deaths deciles 1-10** |
| APR-DRGs | 0.84[4] | 0.20[4] | 0.0-45.0[4] | 0.83[2] | 0.07[2] | 0.0-14.7[2] |
| R-DRGs | 0.80[4] | 0.15[4] | 1.0-42.2[4] | 0.85[2] | 0.08[2] | 0.3-16.2[2] |
| Acuity Index Method (AIM) | | | | 0.78[2] | 0.05[2] | 0.3-12.9[2] |
| Disease Staging | | | | | | |
| Mortality probability | 0.86[4] | 0.27[4] | 0.3-58.4[4] | 0.78[2] | 0.07[2] | 0.8-13.9[2] |
| Stage | 0.79[4] | 0.17[4] | 1.4-49.7[4] | 0.72[2] | 0.02[2] | 0.9-8.8[2] |
| Patient Management Categories | | | | | | |
| Severity Score | 0.82[1,4] | 0.18[1,4] | 0.2-47.3[4] | 0.81[2] | 0.08[2] | 0.5-15.7[2] |
| Resource Intensity Scale | | | | 0.78[2] | 0.08[2] | 0.3-16.2[2] |

| | Pneumonia (N=18,016) | | | Stroke (N=9,407) | | |
|---|---|---|---|---|---|---|
| **Severity Measure** | **C** | **$R^2$** | **% deaths deciles 1-10** | **C** | **$R^2$** | **% deaths deciles 1-10** |
| APR-DRGs | 0.78[3] | 0.10[3] | 0.4-31.1[3] | 0.77[5] | 0.10[5] | 1.2-33.7[5] |
| R-DRGs | 0.83[3] | 0.28[3] | 0.2-42.7[3] | 0.74[5] | 0.07[5] | 2.7-26.5[5] |
| Acuity Index Method (AIM) | 0.73[3] | 0.05[3] | 0.4-24.5[3] | 0.66[5] | 0.03[5] | 2.8-18.2[5] |
| Disease Staging | | | | | | |
| Mortality probability | 0.80[3] | 0.13[3] | 0.2-35.0[3] | 0.74[5] | 0.11[5] | 3.1-33.7[5] |
| Stage | 0.75[3] | 0.08[3] | 0.3-30.1[3] | 0.60[5] | 0.01[5] | 4.8-14.3[5] |
| Patient Management Categories | | | | | | |
| Severity Score | 0.79[3] | 0.11[3] | 0.2-34.3[3] | 0.73[5] | 0.10[5] | 2.9-36.6[5] |
| Resource Intensity Scale | 0.78[3] | 0.10[3] | 0.2-31.8[3] | | | |


1. Iezzoni LI. The risks of risk adjustment. *JAMA.* 1997;278:1600-1607.
2. Landon B, et al. Judging hospitals by severity-adjusted mortality rates: The case of CABG surgery. *Inquiry.* 1996;33:155-166.
3. Iezzoni LI, Shwartz M, Ash AS, Hughes JS, Daley J, Mackiernan YD. Severity measurement methods and judging hospital death rates for pneumonia. *Med Care.* 1996;34:11-28.
4. Iezzoni LI, Ash AS, Shwartz M, Daley J, Hughes JS, Mackiernan YD. Judging hospitals by severity-adjusted mortality rates: The influence of the severity-adjustment method. *Am J Public Health*. 1996;86:1379-1387.
5. Iezzoni LI, Shwartz, Ash AS, Daley J, Mackiernan Y. Using severity-adjusted stroke mortality rates to judge hospitals. *Intl J Quality in Health Care.* 1995;7:81-94.


Comparative performance of severity-adjustment systems based on administrative data, from Iezzoni et al.'s studies of patients with acute myocardial infarction (AMI), coronary artery bypass graft (CABG) surgery, pneumonia, and stroke in the 1992 MedisGroups Comparative Database.

**Results of Evaluation of Risk Adjustment Systems for HCUP Analysis Pertaining to User Criteria**

A limited number of risk adjustment systems met the criteria outlined above for our literature review, though the literature on the performance of many of these systems was rather limited. The additional criteria developed from interviews with potential HCUP users led us to adopt the 3M All-Patient Refined Diagnosis-Related Group (APR-DRG) system with severity and risk-of-mortality classifications for our evaluation of the potential HCUP II QIs and the construction of provider and area estimates. We discuss each criterion in turn.

*"Open" system*. The general DRG methodology and medical terminology is widely understood by hospital administrators and physicians, which facilitates acceptance. The APR-DRG logic is open code, which permits the manual coding and checking of individual medical charts if necessary, and does not depend on specific, "black box" computer software code for implementation. 3M provides extensive user documentation and support, and has working relationships in place with most current and potential HCUP II users.

*Low additional cost*. Data collection for the APR-DRG system is based on standard, widely used abstract systems for individual hospital discharges (specifically, the UB-92). Reliance on currently collected data and information technology systems minimizes additional IT resources and data collection costs. In addition, the system has already been applied to a relatively extensive variety of discharge data, allowing more complete construction of comparative provider and area norms. For many hospitals and states, the incremental data collection costs would be substantial for other established and well-validated severity-adjustment systems that require outpatient data, such as Ambulatory Care Groups (ACG), or medical chart abstraction, such as Atlas™ severity groups. There are many reasons to use these richer adjustment systems when suitable data are available. If the data content of HCUP improves, for example to include additional clinical detail, longitudinal records, and/or ambulatory care information, then more extensive risk adjustment would be feasible.

*One risk adjustment system*. The HCUP II QIs include mortality measures (e.g., AMI mortality rates), complication measures (e.g., diabetes long-term complication rates) and measures of appropriateness (e.g., VBAC rates). Ideally, one would want a risk-adjustment system that includes factors that are suitable for each of these multiple types of performance measures. The current APR-DRG system (along with some competing systems, such as Disease Staging) permits evaluation of both resource use (with a severity-of-illness classification) and outcomes (with a risk-of-mortality classification). In addition, the refinement incorporates a classification system for neonates developed by the National Association of Children's Hospitals and Related Institutions. Therefore the system is theoretically appropriate for the distinctive disease characteristics of patients from all age groups.

*Official recognition.* The APR-DRG system is used by almost all of the state data organizations and health agencies currently reporting comparative hospital data on resource use and mortality. In addition, the Medicare Payment Advisory Commission recently issued a report that advised the adoption of a refined DRG system to improve the accuracy of DRG payments in Medicare, using the APR-DRG system as an example (MEDPAC, 2000). Any such refinement

would very likely to be modeled after the APR-DRG system, given the close relationship with the existing DRG system for Medicare reimbursement and the systems currently use by states and hospitals. HCFA contracts with 3M to maintain the "Grouper" software for the DRG system on an ongoing basis, which would facilitate continuing refinements to the APR-DRG system.

**Conclusion**

As a result of these findings, APR-DRGs were selected for two purposes. The first was to test the impact of risk adjustment on indicators. The second is that APR-DRGs will be implemented in the Version 2 HCUP II QI software, as an optional risk adjustment system. It should be emphasized that other risk-adjustment systems based on ICD-9-CM diagnosis and procedure codes, such as Disease Staging, may work as well or better than APR-DRGs for specific Quality Indicators. Our incorporation of APR-DRGs into the Version 2 software should not be construed as an unequivocal endorsement of this product. Indeed, customized risk-adjustment systems might be more effective than APR-DRGs or any off-the-shelf product, for the reasons outlined on page 148. However, it was beyond the scope of this contract to develop customized risk-adjustment systems for each Quality Indicator. Users may implement other severity stratification systems instead of APR-DRGs if they prefer.

Some indicators could not be risk adjusted using APR-DRGs. Several provider level indicators had denominators that corresponded with an APR-DRG category without severity classifications. Further, all the area indicators only include age-sex risk adjustment, as we did not have information on the APR-DRG distribution of the population (only those discharged from hospitals). It is unclear how risk adjustment with APR-DRGs in these cases (using the discharges in that area) should be interpreted.

# 3.D. General Evidence from the Literature by Indicator Type

The following section contains general evidence reported in the literature regarding indicator types or subgroups. In this report we organize the indicators according to the Donebedian paradigm (Structural, Process, and Outcome measures). Within each type are several subgroups (Volume-Outcome, Utilization, Avoidable hospitalizations or ACSC, and mortality). Indicators within a subgroup are closely related to each other and often have similar concerns, anticipated uses, etc. Often in the literature, indicator subtypes are discussed rather than specific indicators. For instance, and article may discuss mortality indicators, rather than mortality for a specific condition. For some subtypes, such as the ambulatory care sensitive conditions, measure sets have been developed and tested as a set, not as separate indicators. For these reasons, this section outlines the general evidence, concerns, and issues associated with indicators subtype. It is intended that these summaries provide context for the detailed indicator write-ups.

## Structural Measures

Structure describes the setting in which care occurs and the capacity of that setting to produce quality. [22, 50, 210] A substantial literature describes the association with quality of various structural aspects of hospital care - including teaching status, hospital ownership, availability of sophisticated technologies, physician participation on hospital committees, and qualifications of hospital personnel. [203, 211-219] More recent candidate structural measures of hospital quality include patient volumes, the adoption of certain organizational models for inpatient care (e.g., closed ICUs [220] and stroke units [221-223]), and the presence of sophisticated clinical information systems.[224-226] The information contained in the NIS supports consideration of 1) patient volumes, 2) teaching status, and 3) hospital ownership as potential quality indicators for HCUP II. Only patient volume, however, is discussed below, as the former two are not easily changed, and thus would not be good vehicles for quality improvement.

### *Volume-outcome relationships*

An extensive literature indicates that hospitals and physicians with higher patient volumes achieve better outcomes across a broad range of conditions and procedures.[51-53, 56, 58, 61, 64-66, 189, 190, 199, 227-238] The weight of evidence from this extensive literature suggests that, in the absence of accurate or reliable risk adjusted mortality rates, volume data may play important roles in informing patients and purchasers in provider selection.

HCUP I QIs already include in-hospital mortality for many conditions. Volume-based indicators may play a role for conditions for which in-hospital mortality differs significantly from long-term (out-of-hospital) mortality or where adequate risk adjustment requires more clinically detailed information than is available from discharge abstracts.

The literature suggests two basic models for volume-outcome relationships: the intuitive "practice makes perfect" explanation and the less intuitive "selective referral" [52, 230, 239, 240] explanation, which means that high volume does not *cause* high quality, but that a provider is high volume because it *is* a high quality provider (i.e., patients are aware through informal knowledge that superior care available from certain providers, and they go to those providers). The observation that patients often seek their care from hospitals of superior quality even before explicit data such as mortality rates become publicly available [241] supports this explanation.

Studies that have specifically addressed this issue suggest that both explanations play a role, with "practice makes perfect" predominating in emergent conditions, such as acute myocardial infarction [230, 239] and unscheduled cardiac bypass surgery [228], but selective referral accounting for the observed associations for elective bypass surgery [230, 239] and hip replacement [240, 242]. The distinction is relevant to the anticipated response to using a volume-based indicator on quality improvement activities. If selective referral provides the dominant explanation for the relevant volume-outcome relationship, then a volume-based indicator is a 'signal' of a high or low quality provider, and the provider can only improve on the indicator by improving overall quality, communicating that improvement to potential patients, and wait for additional patients to arrive. On the other hand, if higher volume causes better outcomes, then it may not always be the case that directing additional patients to high volume providers is the right policy choice. One study of cardiac surgery indicated improved outcomes as a result of increased patient volumes [233], but a study of pediatric trauma centers corroborated the concern that beyond a certain threshold, increases in volume may strain provider resources and worsen patient outcomes [243]. In reviewing the evidence supporting the volume-based indicators below, we consider the policy implications of diverting patients from local providers to high-volume centers [244] or the possibility of consequent decreases in patient satisfaction [245] the in section on fostering quality improvement (if evidence exists).

In general, one should consider indicators of volume, utilization, and outcomes together. It may be possible that high volume is required for technical skill on a procedure that overall is harmful to patients and prone to overuse. High volume providers may contribute to worse outcomes overall at the population level by exposing more patients to the adverse events and complications associated with the procedure [246]. Small-area research, which has the potential to distinguish between the outcome effects of high-volume and over utilization, sometimes suggests that the health impacts of variations in providers' propensity to subject patients to particular therapies (the utilization effect) dwarf effects of variations in the technical skill required to deliver these therapies (the volume effect), suggesting that utilization, not volume, is a more important determinant of quality.

Carotid endarterectomy (CEA), one of our selected indicators, illustrates the possibility. Publication of favorable clinical trials has led to a dramatic increase in the performance of CEA (see, for example, the North American Symptomatic Carotid Endarterectomy Trial (NASCET) [247] and the Asymptomatic Carotid Atherosclerosis Study (ACAS) [248]).[249-251] In this context, concern has arisen regarding the discrepancy between the efficacy (in a clinical trial) and effectiveness (in the real world) of CEA, especially for patients with asymptomatic carotid stenosis, [252, 253] for which even the case for efficacy is not entirely convincing [254, 255]. Although community practice may replicate the efficacy results of randomized trials in some regions [256] (especially in the case of symptomatic stenosis [257]), outcomes analyses for Medicare patients undergoing CEA indicate a substantial difference between efficacy and effectiveness. [64, 250, 258] Mortality rates among Medicare patients were substantially higher than those reported in the ACAS, even at high volume centers that participated in the trial. [258] One explanation for the discrepancy between the clinical trial results and the observational studies might be the different characteristics of the patient population studied. For example, the ACAS trial excluded all patients over age 80, [248] while 15% of the Medicare patients undergoing CEA fell into this age range. [258] Patients in this age group are higher risk and experience 2-3 times the peri-operative mortality reported for younger patients. [250, 258-260] In summary, high volume might be a quality indicator for only a specific population, and high utilization rates for a broader population may

actually be an indicator of poor quality. Increasing volume by performing more procedures on patients with higher risks of poor outcomes may not be the desirable response (see, for example, a population analysis of outcomes for laminectomy documented worse outcomes for surgeons practicing in areas with a high population rate of procedure performance). [261]

## Process Measures

### *Utilization of Hospital Procedures*

Utilization measures consist of both potentially overused and potentially underused medical procedures. Appendix 8 summarizes studies on appropriateness of procedure utilization. Appropriate use of medical procedures has drawn attention as the cost of health care has increased. Overuse of procedures may increase adverse outcomes by exposing patients to the risk of side effects or complications of treatment,[262-264] and also may lead to unnecessary health care costs.[265] Underuse of procedures may lead to avoidable adverse outcomes.[251, 266, 267] Many different data sources have been used to identify treatments with potential overuse or underuse problems; often, evidence comes from large variations in use across providers or geographic areas.[263, 268] Most attention has been directed to overused surgical procedures,[251, 262-264, 266, 269] because of the availability of reliable information on the performance of these procedures and their substantial costs and associated risks. In recent years, states, insurers, and health plans have all developed initiatives to understand and reduce inappropriate procedure use – including both steps to discourage use of costly "overused" procedures and encourage use of "underused" procedures.[265]

Variations in procedure use across geographic areas and providers may have many explanations. Geographic variations may be due to differences in case-mix, access to care, or different medical "practice patterns" that reflect a combination of uncertainty about the effects of procedures in some patients and provider familiarity with particular methods of treatment.[270, 271] Therefore geographic variation is an indication of possible overuse or underuse of a procedure, but may have other important explanations.

The utilization indicators identified in our review mostly consist of widely used surgical procedures, which at least at the area level are unlikely to vary substantially as a result of small sample sizes. (One exception is incidental appendectomy in the elderly; this procedure has not received much recent attention.)

Each of the procedures identified as potentially overused has an appropriate use in some patients. Appropriateness is influenced by a range of medical considerations including a patient's severity of disease, comorbidities, as well as provider factors such as provider experience. All of these factors may influence the procedure rate that is "appropriate" for a particular hospital. For instance, a hospital may draw more high risk births, leading to a higher cesarean section rate. Appropriate risk-adjustment and other techniques to remove the effects of such hospital-level differences may increase the value of information on hospital procedure rates.

Rates of procedures may also be influenced by non-medical patient characteristics, such as a patient's ability to cope with symptoms or physician practice style[272]. Patient preference may also play a large role in utilization rates. Some patients may be reluctant to undergo any surgical procedure. Other patients may prefer more intensive management, if there is some chance of eliminating disease symptoms.

For all of these reasons, the "ideal" rate for many procedures is not known with certainty. Although *Healthy People 2010* developed target rates for two of the utilization measures – cesarean section and vaginal birth after cesarean section (VBAC) – debate surrounding the

"correct rate" for these and other utilization measures still exists.[273] Much of the evidence on appropriate rates has been developed in a large series of studies by the RAND Corporation, using judgments by expert panels of physicians to rate the appropriateness of procedure use in an extensive set of possible indications, and then to apply these appropriateness criteria to detailed medical chart reviews for a sample of patients undergoing the procedure (to determine overuse) or potentially eligible for the procedure (to determine underuse).[50, 269, 274] Studies assessing the validity of this method have noted moderate agreement about expert judgments between panel members, with kappa scores ranging from .51 for hysterectomy utilization to .83 for underuse of coronary revascularization. The authors concluded that the technique may be useful for evaluating the appropriateness of alternative observed rates of procedure use.[269] However, much of the clinical information included in such indications as those established by the RAND technique are not included in the hospital discharge abstract. In addition, in many cases, the overall rate of procedure use in an area or hospital is not strongly correlated with the share of patients judged to be undergoing the procedure for inappropriate reasons.[268]

These consideration rates suggest that excessive reliance on quality indicators based on utilization rates calculated from hospital discharge abstracts may produce some undesirable effects. Use of potentially underused procedures may increase substantially, but the expanded use might involve patients who are not appropriate candidates.[251, 266] In contrast, use of potentially overused procedures could fall among patients that meet appropriate indications for procedure use. Little evidence exists on the extent to which appropriate and inappropriate use may respond to utilization-based quality indicators, especially indicators based on hospital data.

***Potentially Avoidable Hospital Admissions (ACSC)***

Ambulatory care sensitive conditions are conditions for which hospitalizations may have been avoided through adequate primary care. In general, conditions identified as ambulatory care sensitive conditions have been identified through consensus processes involving panels of expert physicians, using a range of methodologies and decision criteria. Two sets of indicators are widely used. One developed by John Billings in conjunction with the United Hospital Fund of New York includes 28 ambulatory care sensitive conditions, identified by a panel of 6 physicians.[275] The other set was developed by Joel Weissman, and includes 12 avoidable admissions identified through review of the literature review and evaluation by a panel of physicians.[276] These measure sets have been further adapted for pediatric populations.[277, 278] Many of the ambulatory care sensitive conditions have practice guidelines associated with them, including almost all of the chronic conditions and about half of acute medical or pediatric conditions. Most guidelines are consensus statements rather than evidence-based statements, though there are some notable exceptions such as angina. Conditions without explicit guidelines are associated, to varying extents, with ambulatory treatments that are widely viewed as important in preventing hospitalization.

Most of the ambulatory care sensitive conditions involve common types of hospital admissions. However, some are rare admissions, leading to heightened concern about the precision of these measures. For example, among adult medical conditions, admission for dental conditions is rare. Among pediatric conditions, admission for iron deficiency anemia, pediatric mastoiditis, and a primary diagnosis of otitis media (eliminating cases with myriagotomy or tube placement) are rare. Admission for immunization preventable diseases and pediatric urinary tract infections are also somewhat rare.

The extent to which hospitalizations for these conditions are attributable to quality problems and thus the "appropriate" rate of admission for these conditions is not well known. One cause of variation in admission rates is better or worse quality of ambulatory care. In the ambulatory care setting, physicians may fail to prescribe appropriate treatments, follow-up with patients, or offer advice on lifestyle changes. Physicians may fail to comply with guidelines, or may not even be aware of current treatment guidelines.

Another cause is differences in access to care. Several studies have shown that admissions for ambulatory care sensitive conditions are higher in areas with lower socioeconomic status.[275, 279-283] Some investigators have developed evidence to support the view that patients in these areas may not have appropriate access to primary care, due to fewer physicians, lack of insurance, or other difficulties in obtaining care (lack of vacation time to attend doctors appointments, lack of child care, etc.).[284-286] Other causes for variation in admission rates include higher rates of air pollution and other environmental exposures, and higher rates of malnutrition and unhealthy patient behaviors like smoking. One study has indicated that, at least for infants, higher rates of admissions for ambulatory care sensitive conditions is likely to be caused by factors other than genetic predisposition.[287]

Patient health characteristics obviously may influence the rate of hospitalizations. Severity of illness may increase the likelihood that some outpatient treatments may fail. Further, in chronic diseases such as COPD or congestive heart failure, severity of the disease may progress over time. Thus, although good ambulatory care may reduce admissions, as time progresses a patient with one of these chronic illnesses would become more likely to be admitted due to progression of the underlying disease. Similarly, age is often correlated with increased admission rates.[280] Significant comorbid conditions may also complicate outpatient treatment. Some comorbid conditions may increase the likelihood of dangerous complications requiring hospital treatment; other comorbid conditions may be contraindications to the administration of preventive treatments.[280]

Even with appropriate treatment administration and access to care, patient compliance may affect admission rates for ambulatory care sensitive conditions. Several studies have noted that patients frequently do not comply with physician advice on managing their condition. Outpatient therapy for many conditions can be time consuming, costly, or difficult to perform properly, and can have unpleasant side effects. Compliance is a critical and difficult issue in medicine, and may affect admission rates for ambulatory care sensitive conditions.

Finally, differences in threshold for admission may vary between geographic areas, physicians, and even between patients. Some physicians may regularly admit patients with less severe complications, rather than continue to try to manage the condition on an outpatient basis. Even within a physician's own practice, a physician may decide that the patient or caregiver could not handle outpatient treatment due to factors other than medical condition and choose to admit the patient. Home environment, responsibility level of the caregiver, incentives, [288] and other such issues often play into these decisions.

For all of these reasons, the extent to which the reporting of admission rates for ambulatory care sensitive conditions may lead to changes in ambulatory practices and admission rates is unknown. In any case, except for patients who are readmitted soon after a discharge, it is unlikely that the quality of hospital care is a significant determinant of admission rates for ambulatory care sensitive conditions. Rather, ambulatory care sensitive conditions are likely to measure the quality of the health care system as a whole, and especially the quality of primary care, for preventing medical complications. As a result, these measures are likely to be of the

greatest value when calculated at the area level, and when used by public health groups, state data organizations, and other governmental agencies.

## Outcome Measures

### *In-hospital Mortality*

Mortality measures are some of the most widely used quality indicators, and as such they have been subject to the most debate. Mortality measures are used by varying degrees by Hospital Associations, proprietary measures sets (UHC, HCIA, US News and World Report) and have been used by federal agencies (HCFA) and other data projects (AMI mortality in California, CABG mortality in New York CCP).

Large variations in mortality rates have been noted, particularly for nonsurgical conditions. [12, 289] This is likely the result of the "noise" in most mortality measures; mortality for most conditions is relatively rare. Similarly, it has been difficult to determine, with much certainty, whether hospitals with low mortality rates for one condition have low mortality rates for other conditions. One study documented a modest correlation for mortality rates for differing conditions. [37]

The validity of mortality measures as indicators of quality is one of the most widely examined in the literature. Literature reviews have concluded that many mortality rate measures have limited positive predictive value because of noise; that is, highly adjusted mortality rates are likely to occur by chance, so that when they are re-measured in the next year, mortality may no longer appear high. In addition, factors other than hospital quality influence mortality, including patient comorbidities and disease severity. Many studies have concluded that high mortality rates for certain conditions often reflect greater patient disease severity or comorbidity rather than a true quality problem. On the other hand, for some measures, clinically detailed risk adjustment systems do not substantially alter the rates estimated from less detailed discharge data. For a more detailed discussion of risk adjustment issues, see Section 2.D.

Even when mortality measures remain high after risk adjustment, suggesting that a quality problem exists, it is sometimes difficult to identify process-of-care differences that could explain the higher mortality. On the other hand, high outliers may have higher rates of deaths judged "preventable" by experts, though judgments about preventability often have limited inter-rater reliability. Another study noted some correlation between risk-adjusted mortality rates and quality problems identified by peer review organizations. [147]

In general, these studies suggest that noise due to small numbers of deaths is a significant problem for interpreting many mortality measures. Even if estimates are precise enough that noise is unlikely to be a cause of a high apparent mortality rate, differences in disease severity that are difficult to measure may, at least partially, explain the higher rate. However, a number of studies have found that potentially preventable adverse events are at least weakly associated with measured mortality rates.[290, 291]

The impact of mortality reports has also been debated. Consumers may be most interested in mortality measures, and many consumer-aimed quality reports, such as that developed by US News and World Report utilize mortality rates. Physicians appear to be the most skeptical of mortality rates. One study noted a decrease in risk-adjusted mortality rates for all conditions studied after the release of mortality reports in Ohio. [292] However, some have questioned whether this effect represented a true improvement in quality of care, or simply a shifting of deaths to other settings.

## Conclusion

Each of the indicators evaluated in this report belong to one of the subgroups above, and as such, unless otherwise noted, the considerations and evidence discussed above in the corresponding write-up apply to that indicator. For instance, the considerations discussed in the mortality section apply to mortality after CABG and all mortality indicators. When examining the detailed indicator evaluations in the next sections, readers should keep in mind the general evidence discussed above.

# 3.E. Detailed Evidence by Indicator

## Structural Measures

## 3.E.1 Volume Measures

**INDICATOR 1: ABDOMINAL AORTIC ANEURYSM (AAA) REPAIR VOLUME**

| | |
|---|---|
| Indicator | Abdominal aortic aneurysm (AAA) repair, raw volume |
| Relationship to Quality | Better outcomes have been associated with higher volumes. Higher volumes thus represent better quality. |
| Benchmark | Threshold 1: 10 or more procedures per year[195]<br>Threshold 2: 32 or more procedures per year[196, 197] |

Method:

| | |
|---|---|
| Quality Measure | Provider level AAA repair raw volume |
| Outcome of Interest | Discharges with ICD-9 codes 38.34, 38.44, 38.64 in any procedure field and a diagnosis of AAA in any field.<br><br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |
| Population at Risk | Not applicable |

***Evidence from the literature***

**Face validity**. Procedure volume is a surrogate measure of quality; its face validity depends on whether a strong association with outcomes of care is both plausible and widely accepted in the professional community.

Abdominal aortic aneurysm repair requires technical proficiency with the use of complex equipment. Technical errors may lead to clinically significant complications, such as arrhythmias, acute myocardial infarction, colonic ischemia, and death. However, we are not aware of any consensus guidelines or recommendations regarding minimum procedure volume.

**Precision**. Abdominal aortic aneurysmectomy is not as common as the other cardiovascular procedures described in this report; only about 48,600 were performed in the USA in 1997 (1.8 per 10,000 persons).[293] Based on state all-payer databases, the mean annual frequency of abdominal aortic aneurysmectomies was 16.4-18.3 per hospital in Florida (unruptured only) in 1992-1996 [294], 8.4 per hospital in New York (unruptured only) in 1990-1995,[295] and 13.8 per hospital in Maryland in 1990-1995.[296]

The number of abdominal aortic aneurysm resections is measured accurately with discharge data; in fact, discharge data are probably the best available source for hospital volume information. However, the relatively small number of procedures performed annually at most hospitals suggests that annual volume may be subject to considerable random variation.

**Minimum bias**. Volume measures are not subject to bias due to disease severity and comorbidities. For this reason, risk-adjustment is not appropriate. Although volume measures

are theoretically subject to bias due to variation across hospitals in the use of outpatient surgery facilities, less than 1% of AAA repairs in 1996 were performed in ambulatory settings." [297]

**Construct validity**. Volume is not a direct measure of the quality or outcomes of care. Although higher volumes have been repeatedly associated with better outcomes after abdominal aortic aneurysm resection, these findings may be limited by inadequate risk adjustment.

All but one of 15 studies published since 1985 demonstrated a significant association between either hospital or surgeon volume and mortality after abdominal aortic aneurysm repair. However, three sets of these studies (e.g., Hannan 1989 and 1992[53, 195]; Kazmers, 1996 and Khuri, 1999[196, 298]; Pronovost, 1999 and Dardik 1999 and 1998 [62, 197, 296]) appear to include overlapping cases, which may exaggerate the consistency of their results. Two studies of intact aneurysms found that hospital and surgeon volume were significant independent predictors of inpatient mortality, adjusting for other hospital and patient characteristics,[294, 296] whereas two studies (one of which included ruptured aneurysms) found significant effects only for hospital volume.[195, 197] The three studies that focused exclusively on ruptured aneurysms reported a significant effect only for surgeon volume.[62, 195, 299] Six studies that considered only hospital volume found lower mortality at high-volume hospitals;[188, 196, 295, 300] two of these studies found a significant association among intact aneurysms but not among ruptured aneurysms.[301, 302]

The only completely negative study was Khuri and colleagues'[298] evaluation of abdominal aortic aneurysmectomies performed in Veterans Affairs hospitals from 1991 through 1997. Their study was the only one that used clinical data, which allowed them to construct a relatively powerful risk-adjustment model. Yet they found only a very weak, nonsignificant association between procedure or specialty volume and risk-adjusted 30-day mortality.

Several studies of this topic have explored whether the volume of something other than the index procedure may be a more powerful predictor of mortality than the volume of the index procedure. The underlying hypothesis is that experience acquired on related, but not identical, cases may lead to improved outcomes. However, two studies failed to show any significant association between the *total* physician[299] or hospital[62] volume of abdominal aortic aneurysmectomies and mortality among *ruptured* aneurysms, whereas one study showed similar associations between total hospital volume and mortality for both ruptured and intact aneurysms.[188] Of two studies that evaluated the impact of total vascular surgery volume, one found a significant effect on mortality for both ruptured and intact aneurysms,[302] whereas the other found no effect for intact aneurysms (adjusting for procedure-specific volume).[298] One study found that the hospital volume of surgery for ruptured aneurysms was not associated with postoperative inpatient mortality, but it was associated with fewer inpatient deaths for ruptured aneurysms, suggesting that high-volume hospitals may manage ruptured aneurysms more aggressively.[303]

It is difficult to determine an appropriate hospital volume threshold from the published literature, because numerous studies reported the volume effect only in linear terms[294, 299] or used volume categories simply to display unadjusted mortality.[196, 295, 300-302] Among the studies that either reported indirectly standardized mortality rates by volume strata, or used volume categories in multi-level regression models, the recommended hospital volume thresholds were 10,[195] 20,[188] or 36[197] cases per year of either intact or ruptured aneurysms, or 50 elective repairs of intact aneurysms per 6 years.[296] Among the studies that analyzed volume as a linear effect, but displayed crude mortality by stratum, the stratum definitions were 10 or 21 intact aneurysms per year,[300, 301] 32 intact aneurysms per 3 years,[196] or 100 cases of any vascular surgery per year.[302]

Although volume-outcome associations have been demonstrated for abdominal aortic aneurysmectomy, volume seems likely to be both insensitive and nonspecific as a measure of quality. Nonetheless, it has been estimated that shifting patients in California from low-volume to high-volume hospitals would avert 40 deaths per year, given that 64% of all operations are performed in low-volume hospitals.[190]

**Fosters true quality improvement**. One possible adverse effect of volume-based measures is to encourage low-volume providers (who may also provide poorer quality of care) to increase their volume, simply to reach an artificial threshold. Such responses would probably not improve patient outcomes to the same extent as moving patients from low-volume to high-volume hospitals. For example, Hannan and colleagues [195] found that the 22 surgeons in New York who increased their volume from 1-6 aneurysm repairs in 1982-84 to 14 or more aneurysm repairs in 1985-87 achieved a minimal decrease in standardized mortality, from 6.8% to 6.2%. The subset of these surgeons who increased their volume to at least 22 cases achieved an important but nonsignificant decrease in standardized mortality, from 5.8% to 2.5%. At the extreme, hospitals may loosen eligibility criteria and perform procedures on patients who are marginal or inappropriate candidates. These arguments would not apply to the subset of ruptured aneurysms, which are not subject to volume manipulation. However, shutting down low-volume hospitals and transferring procedures to high-volume hospitals may worsen outcomes for ruptured aneurysms by delaying surgical intervention.

**Prior use.** Abdominal aortic aneurysmectomy volume has not been widely used as an indicator of quality. In its Web site, the Pacific Business Group on Health[304] states that "one marker of how well a hospital is likely to perform is the experience of the hospital and its surgical team...in the absence of data to compare hospitals on their complications and survival rates, you can begin evaluating experience by looking at the number of (abdominal aortic aneurysm) surgeries a hospital performs each year." The Center for Medical Consumers posts hospital-specific and operator-specific volumes of "resection of aorta with replacement" for New York hospitals.[305]

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Procedure volume* | | |
| Raw mean volume (standard deviation) | 14 / (16) | |
| Median/90th/95th percentile | 8 / 36 / 46 | |
| Stability over time, mean in 1995 / mean in 1997 | 14 / 14 | Stable |
| *Percentage of procedures at high volume hospitals* | | *Mod / Low* |
| Percentage threshold 1 (% hosp at threshold) | 84% (44%) | |
| Percentage threshold 2 (% hosp at threshold) | 43% (12%) | |
| *Persistence of high volume* | | *Moderate* |
| High volume remaining high, 95/96 (Threshold 1, threshold 2) | 86% / 81% | |
| High volume remaining high, 96/97 (Threshold 1, threshold 2) | 87% / 76% | |

**Procedure Volume.** In 1996, 727 hospitals (54.1% of providers) performed at least one procedure. Of these hospitals, the mean (standard deviation) of the number of procedures was 14 (16). The median was 8, and the 90th and 95th percentile was 36 and 46, respectively. In general,

there are many hospitals with lower volumes and a few hospitals with much higher volumes. Overall procedure volume was stable over the 1995-1997 time period. The mean number of procedures performed in 1995 and 1997 was 14.

**Percentage of Procedures at High Volume Hospitals.** A moderate to low percentage of procedures were performed at high volume hospitals, depending on which threshold you use. At the threshold 1, 83.9% of AAA repair procedures were performed at 'high volume' providers (and 44.3% of providers are 'high volume'). At the threshold 2, 43.0% were performed at 'high volume' providers (and 12.2% of providers are 'high volume').

**Persistence of High Volume**. High volume status was moderately persistent over time, depending on the volume threshold used. At threshold 1, 86.2% of high volume providers in 1995 were also high volume in 1996. Similarly, 81.1% of high volume providers in 1996 were also high volume in 1997. At threshold 2, 87.0% of high volume providers in 1995 were also high volume in 1996. Similarly, 75.9% of high volume providers in 1996 were also high volume in 1997.

**Construct validity**. We estimated the correlation between AAA volume and mortality, adjusting for patient characteristics such as age, sex, and APR-DRG. Volume for carotid endarterectomy is independently and negatively correlated with mortality for carotid endarterectomy ($r=-.35$, $p<.001$).

***Discussion***

AAA repair is a relatively rare procedure. Our empirical analysis found a mean of 14 procedures per year. While a large number of hospitals perform at least one procedure, only 44% (threshold 1) or 12% (threshold 2) of hospitals are actually high volume. The relationship between volume and outcome has been established in the literature, specifically that higher volume hospitals have lower mortality than lower volume hospitals; differences in patient case-mix do not account fully for these relationships. Our empirical analysis noted that AAA repair volume was negatively correlated to AAA repair mortality. However, our results do not include the complex risk adjustment contained in the studies reported in the literature.

This indicator is measured with great precision, as is expected with all volume indicators. It is expected that volume for AAA repair would be measured precisely using discharge abstract data. Most procedures are performed in an inpatient setting.

The volume-outcome relationship on which this indicator is based may not hold over time, as providers become more experienced or as technology changes. It is important then to revisit the volume-outcome relationship to ensure the validity of this indicator. Overall, volume measures are not direct measures of quality, and are relatively insensitive. For this reason they should be used with caution and in conjunction with other measure of mortality, to ensure that increasing volumes truly improve patient outcome.

Our empirical analyses found that most AAA procedures are not actually performed at high volume hospitals already. This leaves ample room for improvement. Relatively few hospitals are high volume. It is unclear whether simply increasing volume at low volume hospitals would actually improve outcomes. It is possible that hospitals could increase volume simply by increasing the number of borderline or inappropriate procedures performed. For this reason, it may be prudent to examine this indicator alongside area rates for this procedure, and examinations of the appropriateness of the procedures.

Performance for this indicator is highly dependent on the volume threshold used.

Overall, this indicator is recommended for inclusion in the HCUP II QI set. Specific caveats should be kept in mind when using this indicator. As a volume indicator, this indicator is a proxy measure for quality, and it is recommended that it be used with other indicators. Theoretically, providers could increase the number of procedures on patients with questionable indications as well, though this is more difficult than for other indicators.

**INDICATOR 2: CAROTID ENDARTERECTOMY (CE) VOLUME**

| | |
|---|---|
| Indicator | Carotid endarterectomy raw volume |
| Relationship to Quality | Better outcomes have been associated with higher volumes. Higher volumes thus represent better quality. |
| Benchmark | Threshold 1: 50 or more procedures per year[188]<br>Threshold 2: 101 or more procedures per year[189, 190] |

Method:

| | |
|---|---|
| Quality Measure | Provider level CE raw volume |
| Outcome of Interest | Discharges with ICD-9 Code 38.12 in any procedure field.<br><br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |
| Population at Risk | Not applicable |

***Evidence from the literature***

**Face validity**. Procedure volume is a surrogate measure of quality; its face validity depends on whether a strong association with outcomes of care is both plausible and widely accepted in the professional community.

Carotid endarterectomy (CEA) is a procedure that requires technical proficiency with the use of complex equipment. Technical errors may lead to clinically significant complications, such as abrupt carotid occlusion with or without stroke, myocardial infarction, and death. In two major randomized trials of CEA, researchers pre-selected centers with low rates of perioperative stroke and death, in the belief that these measures reflect surgical skill and quality of care.[306, 307] As a result, recent professional guidelines focus on the importance of monitoring surgical outcomes, and the avoidance of promoting volume standards.[308]

**Precision**. Publication of the North American Symptomatic Carotid Endarterectomy Trial (NASCET) [247] and the Asymptomatic Carotid Atherosclerosis Study (ACAS) [248] has led to a dramatic increase in the performance of CEA.[249-251] Approximately 144,000 CEAs were performed in the USA in 1997 (5.3 per 10,000 persons).[309] However, many hospitals perform relatively few procedures, suggesting that the actual annual count of procedures may not be a reliable guide to the number of procedures performed on an ongoing basis. For example, 60% of institutions performed fewer than 17 procedures per year in one study of Medicare beneficiaries.[258] Approximately 50% of CEAs performed on Medicare beneficiaries in one state occurred in hospitals performing 21 or fewer operations annually.[64] Although these numbers involve Medicare patients only, Medicare beneficiaries comprise approximately 70% of the patients who undergo CEA, so total patient volumes are unlikely to differ substantially.

The number of CEA procedures is measured accurately with discharge data; in fact, discharge data are probably the best available source for hospital volume information.

**Minimum bias**. Volume measures are not subject to bias due to disease severity and comorbidities. For this reason, risk-adjustment is not appropriate. Although volume measures are theoretically subject to bias due to variation across hospitals in the use of outpatient surgery facilities, less than 1% of CEA surgeries in 1996 were performed in ambulatory settings." [297]

**Construct validity**. Volume is not a direct measure of the quality or outcomes of care. Although higher volumes have been repeatedly associated with better outcomes after CEA, these findings may be limited by inadequate risk adjustment. Only two studies outside the Veterans Affairs system were based on clinical data sets that included "indication for surgery"; both used hospital-specific Medicare volume, not total volume, as the key independent variable.[64, 310]

Surgeons and hospitals with higher patient volumes tend to have fewer adverse outcomes, including new strokes and deaths.[64, 258, 310-312] Only two major studies, one limited to Veterans Affairs medical centers[298] and the other from Finland,[313] failed to show a significant hospital volume-outcome relationship for CEA. The magnitude of this effect was impressive in the two studies with the best severity adjustment; for example, Cebul et al. found that undergoing surgery in a high-volume hospital was associated with a 71% reduction in the risk of stroke or death at 30 days, after adjustment for age, gender, indication for surgery, renal insufficiency, and two cardiovascular comorbidities.[64, 310] In the study by Karp et al., the risk of severe stroke or death was 2.6 times higher at the lowest-volume hospitals than at the highest-volume hospitals.[64, 310] Other studies with more limited risk adjustment have reported adjusted odds ratios at low-volume hospitals of 1.28 for death [189] and 2.5-3.1 for "death or definite stroke" (based on ICD-9-CM codes).[64, 310]

Optimal volume thresholds are difficult to determine, because most studies have only counted Medicare cases. Using a statewide database in New York, Hannan et al. found that hospitals with fewer than 101 cases per year had elevated risk-adjusted mortality. Researchers using Medicare data have reported volume threshold effects at 40[64, 310] and 62[64, 310] cases per year, or continuous effects over a volume range from <7 or <11 up to >21 or >50 cases per year.[64, 310]

**Fosters true quality improvement**. One possible adverse effect of volume-based measures is to encourage low-volume providers (who may also provide poorer quality of care) to increase their volume, perhaps to reach a threshold of 101 cases per year. Such responses would probably not improve patient outcomes to the same extent as moving patients from low-volume to high-volume hospitals. Indeed, hospitals may loosen eligibility criteria and perform procedures in patients who are marginal or inappropriate candidates. This possibility is worrisome, because the benefits of CEA in asymptomatic patients are modest and easily outweighed by high postoperative complication rates.[253, 254, 258, 314] Outcomes analyses among Medicare patients undergoing CEA indicate a substantial difference between efficacy and effectiveness.[64, 250, 258]. Mortality rates among Medicare patients were substantially higher than those reported in the ACAS, even at high-volume centers that participated in the trial,[258] because of more liberal patient selection. For example, the ACAS excluded all patients over 80 years of age,[248] but 15% of the Medicare patients undergoing CEA outside ACAS were in this age range.[258] Patients over 80 years of age experience 2-3 times the perioperative mortality reported for younger patients.[250, 258-260]

Despite this caveat, previous studies have shown either no relationship between provider volume and patient selection[258, 298] or a tendency for high volume providers to operate on sicker patients.[64] To address this issue, one would ideally consider provider volume in conjunction with the distribution of indications for surgery or major comorbidities. Unfortunately, the indication for surgery is not obtainable from the HCUP database.

The alternative of shutting down low-volume hospitals and transferring procedures to high-volume hospitals may overload these providers and impair access to care.

**Prior use.** CEA volume has not been widely used as an indicator of quality, although it has been advocated as such. In its Web site, the Pacific Business Group on Health[304] states that "one marker of how well a hospital is likely to perform is the experience of the hospital and its surgical team...in the absence of data to compare hospitals on their complications and survival rates, you can begin evaluating experience by looking at the number of (CEA) surgeries a hospital performs each year." The Center for Medical Consumers posts hospital-specific and operator-specific CEA volumes for New York hospitals.[305]

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Procedure volume* | | |
| Raw mean volume (standard deviation) | 52 (60) | |
| Median/90th/95th percentile | 31.5 / 129 / 169 | |
| Stability over time, mean in 1995 / mean in 1997 | 52 / 54 | Stable |
| *Percentage of procedures at high volume hospitals* | | *Moderate* |
| Percentage threshold 1 (% hosp at threshold) | 78% (37%) | |
| Percentage threshold 2 (% hosp at threshold) | 51% (17%) | |
| *Persistence of high volume* | | *High/ Mod* |
| High volume remaining high, 95/96 (Threshold 1, threshold 2) | 94% / 88% | |
| High volume remaining high, 96/97 (Threshold 1, threshold 2) | 90% / 88% | |

**Procedure Volume.** In 1996, 904 hospitals (67.2% of providers) performed at least one procedure. Of these hospitals, the mean (standard deviation) of the number of procedures was 52 (60). The median was 31.5, and the 90th and 95th percentile was 129 and 169, respectively. In general, there are many hospitals with lower volumes and a few hospitals with much higher volumes. Overall procedure volume was stable over the 1995-1997 time period. The mean number of procedures performed in 1995 and 1997 was 52 and 54, respectively.

**Percentage of Procedures at High Volume Hospitals.** A moderate percentage of procedures were performed at high volume hospitals. At the threshold 1, 77.8% of carotid endarterectomy procedures were performed at 'high volume' providers (and 37% of providers are 'high volume'). At the threshold 2, 51.0% were performed at 'high volume' providers (and 17% of providers are 'high volume').

**Persistence of High Volume**. High volume status was highly persistent over time. At threshold 1, 93.5% of high volume providers in 1995 were also high volume in 1996. Similarly, 89.7% of high volume providers in 1996 were also high volume in 1997. At threshold 2, 87.5% of high volume providers in 1995 were also high volume in 1996. Similarly, 87.8% of high volume providers in 1996 were also high volume in 1997.

**Construct validity**. As we did not retain the carotid endarterectomy mortality indicator due to inadequate precision we were not able to test the construct validity of this indicator. However, CE volume is negatively correlated with several other mortality indicators: CABG ($r=-.26$, $p<.0001$), AAA repair ($r=-.38$, $p<.0001$) and craniotomy ($r=-.18$, $p<.0001$).

***Discussion***

Carotid endarterectomy is a fairly common procedure. Our empirical analysis found a mean of 52 procedures per year. While a large number of hospitals perform at least one procedure, only 37% (threshold 1) or 17% (threshold 2) of hospitals are actually high volume. The relationship between volume and outcome has been established in the literature, specifically that higher volume hospitals have lower mortality and post-operative stroke rates than lower volume hospitals; differences in patient case-mix do not account fully for these relationships. The relationship may be stronger for hospital volume than for operator volume. Nonetheless, providers may want to examine operator volume as well as hospital volume.

This indicator is measured with great precision, as is expected with all volume indicators. It is expected that volume for carotid endarterectomy would be measured precisely using discharge abstract data. Most procedures are performed in an inpatient setting.

The volume-outcome relationship on which this indicator is based may not hold over time, as providers become more experienced or as technology changes. It is important then to revisit the volume-outcome relationship to ensure the validity of this indicator. Overall, volume measures are not direct measures of quality, and are relatively insensitive. For this reason they should be used with caution and in conjunction with other measure of mortality, to ensure that increasing volumes truly improve patient outcomes.

Our empirical analyses found that many CE procedures are actually performed at high volume hospitals already. This leaves some room for improvement, but not as much as for other indicators. However, relatively few hospitals are high volume. It is unclear whether simply increasing volume at low volume hospitals would actually improve outcomes. It is possible that hospitals could increase volume simply by increasing the number of borderline or inappropriate procedures performed. For this reason, it may be prudent to examine this indicator alongside area rates for this procedure, and examinations of the appropriateness of the procedures.

Overall, this indicator is recommended for inclusion in the HCUP II QI set. Specific caveats should be kept in mind when using this indicator. As a volume indicator, this indicator is a proxy measure for quality, and it is recommended that it be used with other indicators. Providers could increase the number of procedures on patients with questionable indications as well.

## INDICATOR 3: CORONARY ARTERY BYPASS GRAFT (CABG) VOLUME

| | |
|---|---|
| Indicator | Coronary artery bypass graft (CABG) raw volume |
| Relationship to Quality | Better outcomes have been associated with higher volumes. Higher volumes thus represent better quality. |
| Benchmark | Threshold 1: 100 or more procedures per year[193]<br>Threshold 2: 200 or more procedures per year[54] |

Method:

| | |
|---|---|
| Quality Measure | Provider level CABG raw volume |
| Outcome of Interest | Discharges with ICD-9 Codes 36.10 – 36.19 in any procedure field.<br><br>Age 40 years and older.<br><br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |
| Population at Risk | Not applicable |

***Evidence from the literature***

**Face validity**. Procedure volume is a surrogate measure of quality; its face validity depends on whether a strong association with outcomes of care is both plausible and widely accepted in the professional community.

CABG is a procedure that requires technical proficiency with the use of complex equipment. Technical errors may lead to clinically significant complications, such as myocardial infarction, stroke, and death. On the basis of this knowledge and empirical literature (summarized below), the American Heart Association (AHA) and the American College of Cardiology (ACC) have argued for "careful outcome tracking" and supported "monitoring institutions or individuals who annually perform <100 cases." Noting that "some institutions and practitioners maintain excellent outcomes despite relatively low volumes," this panel concluded that "credentialing policies based on conclusions drawn from these data must be made with caution."[193] A committee of the Society of Thoracic Surgeons reaffirmed that "until conclusive data become available that link volume to outcome, volume should not be used as a criterion for credentialing of cardiac surgeons…each surgeon should be evaluated on his or her individual results."[315]

**Precision**. The frequency of CABG has been relatively stable over the past 15 years, as percutaneous coronary interventions have become more popular. Approximately 366,000 CABG were performed in the USA in 1997 (13.5 per 10,000 persons).[316] Several states, including New York, New Jersey, Pennsylvania, and California, have started statewide programs to monitor the frequency and clinical outcomes of CABG surgery. These states differ markedly in mean CABG volume: 627 (range 94-1,814) in New York (1996); 636 (range 111-1,119) in New Jersey (1996-97, annualized); 449 (range 121-1,165) in Pennsylvania (1994-95, annualized); and 266 (range 3-1,643) in California (1996). [317, 318]

The number of CABG procedures is measured accurately with discharge data; in fact, discharge data are probably the best available source for hospital volume information. The large

number of procedures performed annually at most hospitals suggests that annual volume is not subject to considerable random variation (except perhaps in California and similar states). Indeed, Hannan et al. reported year to year hospital volume correlations of 0.96-0.97 in New York.[118]

**Minimum bias**. Volume measures are not subject to bias due to disease severity and comorbidities. For this reason, risk-adjustment is not appropriate. Although volume measures are theoretically subject to bias due to variation across hospitals in the use of outpatient surgery facilities, less than 1% of CABG surgeries in 1996 were performed in ambulatory settings.[297]

**Construct validity**. Volume is not a direct measure of the quality or outcomes of care. Although higher volumes have been repeatedly associated with better outcomes after CABG, these findings may be limited by inadequate risk adjustment. Sowden et al. [319] systematically reviewed 15 studies of the volume-outcome relationship for CABG; six used non-overlapping data and reported effect estimates for fixed volume categories. Among these six studies, the apparent benefit of high CABG volume (>200 cases per year) diminished as casemix adjustment improved. Because casemix adjustment was generally more complete in more recent studies, the authors could not exclude the possibility that the benefit of high volume actually decreased between 1972 and 1991.

Using a comprehensive clinical database to adjust for age, gender, unstable angina, ejection fraction, functional class, shock, preoperative intra-aortic balloon pump, recent myocardial infarction, and several comorbidities, Hannan found that the adjusted relative risk of inpatient death at high-volume hospitals (>200 cases per year) in 1989-1992 was 0.84, compared with low-volume hospitals.[118] However, only 3.3% of patients in that study underwent CABG at a low-volume hospital. Another recent study, based on a clinical dataset from the Department of Veterans Affairs, reported a very similar adjusted relative risk of 1.33 at hospitals with less than 101 CABG per year, compared with higher-volume centers.[320]

Older studies using hospital discharge data found larger effects of hospital volume. Differences in risk-adjusted mortality rates across volume quartiles (20-100, 101-200, 201-350, >350 cases per year) were larger for non-scheduled operations (7.7%, 5.5%, 5.9%, and 4.6%, respectively) than for scheduled operations (3.0%, 2.7%, 2.9%, and 2.2%, respectively) in one study.[228] Analyses using instrumental variables suggested that much, if not all, of the volume effect may be due to "selective referral" of patients to high-quality centers.[230, 239] Of course, the direction of causation (e.g., higher volume leads to better outcomes, or vice versa) may not affect the validity of using hospital volume as a marker of quality.

Studies of surgeon volume are less directly relevant to the HCUP Quality Indicator project, but one recent study (from New York) demonstrated a statistically significant association between surgeon volume and mortality, which appears to be decreasing over time.[233] Specifically, surgeons who performed 50 or fewer CABG in 1989 had a risk-adjusted mortality rate 2.2 times greater than that of surgeons who performed 150 or more CABG. This ratio decreased to 1.89 in 1990, 1.39 in 1991, and 1.36 in 1992. Only in 1989 and 1990 were the mortality differences across surgeon volume categories statistically significant. Two earlier studies of surgeon volume generated counter-intuitive and difficult-to-interpret results.[227, 321]

Although volume-outcome associations have been demonstrated for CABG, volume seems likely to both insensitive and nonspecific as a measure of quality. For example, Hannan found that some low-volume surgeons achieved outstanding risk-adjusted mortality rates of 2.1% or less in 1992; these surgeons were either transiently low-volume or new to New York State.

Nonetheless, it has been estimated that shifting patients in California from low-volume to high-volume hospitals would avert 258 deaths per year.[190]

**Fosters true quality improvement**. One possible adverse effect of volume-based measures is to encourage low-volume providers (who may also provide poorer quality of care) to increase their volume, simply to reach a threshold of 200 cases per year. Such responses would probably not improve patient outcomes to the same extent as moving patients from low-volume to high-volume hospitals. At the extreme, hospitals may loosen eligibility criteria and perform procedures on patients who are marginal or inappropriate candidates. The alternative of shutting down low-volume hospitals and transferring procedures to high-volume hospitals may overload these providers and impair access to care.

**Prior use.** CABG volume has not been widely used as an indicator of quality, although specific volume thresholds have been suggested as "standards" for the profession. In its Web site, the Pacific Business Group on Health [304] states that "one marker of how well a hospital is likely to perform is the experience of the hospital and its surgical team...in the absence of data to compare hospitals on their complications and survival rates, you can begin evaluating experience by looking at the number of (CABG) surgeries a hospital performs each year."

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Procedure volume* | | |
| Raw mean volume/standard deviation | 399 (338) | |
| Median/90th/95th percentile | 293 / 830 / 1095 | |
| Stability over time, mean in 1995 / mean in 1997 | 375 / 401 | Increasing |
| *Percentage of procedures at high volume hospitals* | | *High* |
| Percentage threshold 1 (% hosp at threshold) | 98% (88%) | |
| Percentage threshold 2 (% hosp at threshold) | 91% (68%) | |
| *Persistence of high volume* | | *High* |
| High volume remaining high, 95/96 (Threshold 1, threshold 2) | 99% / 97% | |
| High volume remaining high, 96/97 (Threshold 1, threshold 2) | 98% / 97% | |

**Procedure Volume**. In 1996, 307 hospitals (26.4% of providers) performed at least one procedure. Of these hospitals, the mean (standard deviation) of the number of procedures was 399 (338). The median was 293, and the 90th and 95th percentile was 830 and 1095, respectively. In general, there are a moderate number of hospitals with lower volumes and a few hospitals with much higher volumes. Overall procedure volume grew over the 1995-1997 time period. The mean number of procedures performed in 1995 and 1997 was 375 and 401, respectively.

**Percentage of Procedures at High Volume Hospitals.** A high percentage of procedures were performed at high volume hospitals. At the threshold 1, 98.3% CABG procedures were performed at 'high volume' providers (and 88% of providers are 'high volume'). At the threshold 2, 90.7% were performed at 'high volume' providers (and 68% of providers are 'high volume').

**Persistence of High Volume.** High volume status was highly persistent over time. At threshold 1, 99.0% of high volume providers in 1995 were also high volume in 1996. Similarly, 98.1% of high volume providers in 1996 were also high volume in 1997. At threshold 2, 96.9%

of high volume providers in 1995 were also high volume in 1996. Similarly, 97.5% of high volume providers in 1996 were also high volume in 1997.

**Construct validity.** We estimated the correlation between CABG volume and mortality, adjusting for patient characteristics such as age, sex, and APR-DRG. Volume for CABG is independently and negatively correlated with mortality for CABG ($r=-.29$, $p<.001$).

***Discussion***

CABG is a technically difficult, yet relatively common procedure. Our empirical analysis found a mean of 399 procedures per year. A substantial number of hospitals perform at least one procedure, and 88% (threshold 1) or 68% (threshold 2) of hospitals are actually high volume. Higher volumes of coronary artery bypass graft (CABG) have been associated with better outcomes, namely fewer deaths. While several studies have demonstrated this relationship, these studies also have some flaws. Differences in case mix and the extremely low proportion of procedures taking place in low volume hospitals may account for some of the differences between high volume and low volume hospital. The AHA/ACC has recommended that since some low volume hospitals have very good outcomes, other outcomes measures besides volume should be used to evaluate individual surgeons performance. However, the AHA/ACC does not make a recommendation based on volume as a quality indicator. Providers may want to examine operator volume as well as hospital volume. Our empirical analysis noted that CABG volume was very slightly negatively correlated to CABG mortality. However, our results do not include the complex risk adjustment contained in the studies reported in the literature; this could explain the lack of strong association seen in our results.

This indicator is measured with great precision, as is expected with all volume indicators. It is expected that volume for CABG would be measured precisely using discharge abstract data. Most procedures are performed in an inpatient setting.

The volume-outcome relationship on which this indicator is based may not hold over time, as providers become more experienced or as technology changes. It is important then to revisit the volume-outcome relationship to ensure the validity of this indicator. Overall, volume measures are not direct measures of quality, and are relatively insensitive. For this reason they should be used with caution and in conjunction with measures of mortality and of quality of care within that field (in this case cardiac care), to ensure that increasing volumes truly improve patient outcomes.

Our empirical analyses found that most CABG procedures are actually performed at high volume hospitals already. This leaves little room for improvement. Further, most hospitals are high volume. It is unclear whether simply increasing volume at the few remaining low volume hospitals would actually improve outcomes. It is possible that hospitals could increase volume simply by increasing the number of borderline or inappropriate procedures performed. For this reason, it may be prudent to examine this indicator alongside area rates for this procedure, and examinations of the appropriateness of the procedures.

Overall, this indicator is recommended for inclusion in the HCUP II QI set. Specific caveats should be kept in mind when using this indicator. As a volume indicator, this indicator is a proxy measure for quality, and it is recommended that it be used with other indicators. Theoretically, providers could increase the number of procedures on patients with questionable indications as well, though this is more difficult than for other indicators.

## INDICATOR 4: ESOPHAGEAL RESECTION VOLUME

| | |
|---|---|
| Indicator | Esophageal resection raw volume |
| Relationship to Quality | Better outcomes have been associated with higher volumes. Higher volumes thus represent better quality. |
| Benchmark | Threshold 1: 6 or more procedures per year[198]<br>Threshold 2: 7 or more procedures per year[190, 198] |

Method:

| | |
|---|---|
| Quality Measure | Provider level esophageal resection raw volume |
| Outcome of Interest | Discharges with ICD-9 Codes 42.40 – 42.42 in any procedure field and diagnosis code of esophageal cancer in any field.<br><br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |
| Population at Risk | Not applicable |

***Evidence from the literature***

**Face validity**. Procedure volume is a surrogate measure of quality; its face validity depends on whether a strong association with outcomes of care is both plausible and widely accepted in the professional community.

Esophageal cancer surgery requires technical proficiency; errors in surgical technique or management may lead to clinically significant complications, such as sepsis, pneumonia, anastomotic breakdown, and death. However, we are not aware of any consensus guidelines or recommendations regarding minimum procedure volume. The National Cancer Policy Board of the Institute of Medicine and the National Research Council recommends that cancer "patients undergoing procedures that are technically difficult to perform and have been associated with higher mortality in lower-volume settings (including esophagectomy) receive care at facilities with extensive experience (e.g., high-volume facilities)."

**Precision**. The number of esophagectomies is measured accurately with discharge data; in fact, discharge data are probably the best available source for hospital volume information. Although a few facilities have relatively high volumes, most (e.g., 239 of 273 California hospitals)[198] perform 10 or fewer esophagectomies for cancer during a **5-year** period. As a result, this measure is expected to have poor precision.

**Minimum bias**. Volume measures are not subject to bias due to disease severity and comorbidities. For this reason, risk-adjustment is not appropriate. Although volume measures are theoretically subject to bias due to variation across hospitals in the use of outpatient surgery facilities, less than 1% of resections in 1996 were performed in ambulatory settings." [297]

**Construct validity**. Volume is not a direct measure of the quality or outcomes of care. Although higher volumes have been repeatedly associated with better outcomes after esophageal surgery, these findings may be limited by inadequate risk adjustment.

Only one study used clinical data to estimate the association between hospital volume and mortality following esophageal cancer surgery. Begg et al.[65] analyzed retrospective cohort data from the Surveillance, Epidemiology, and End Results(SEER)-Medicare linked database

from 1984 through 1993. The crude 30-day mortality rate was 17.3% at hospitals that performed 1-5 esophagectomies on Medicare patients during the study period, versus 3.9% and 3.4% at hospitals that performed 6-10 and 11 or more esophagectomies, respectively. The association between volume and mortality remained highly significant (p<.001) in a multivariate model, adjusting for the number of comorbidities, cancer stage and volume, and age.

Two other studies using hospital discharge data found similar effects of hospital volume. Using 1990-94 data from California, Patti et al.[198] estimated risk-adjusted mortality rates of 17%, 19%, 10%, 16%, and 6% across five hospital volume categories (e.g., 1-5, 6-10, 11-20, 21-30, and >30 procedures during the 5-year study period). Their risk adjustment was quite limited; only the year of surgery, age, sex, race, payer source, tumor location, and the total number of secondary diagnoses were included. Using 1990-97 data from Maryland (adjusting only for age and payer source), Gordon et al.[322] estimated that the adjusted odds of death at minimal-volume (<11 "complex gastrointestinal procedures" per year) and low-volume (11-20 procedures/year) hospitals were 3.8 and 4.0 times that at a high-volume hospital (214 procedures/year). However, the generalizability of these results is limited by the fact that the last category included only one hospital. An older British study found a surgeon volume effect, but did not consider hospital volume.[323]

Although volume-outcome associations have been demonstrated for esophageal cancer surgery, volume seems likely to both insensitive and nonspecific as a measure of quality. It has been estimated that shifting patients in California from low-volume to high-volume hospitals would avert only 7 deaths per year, although 77% of all operations are performed in low-volume hospitals.[190]

**Fosters true quality improvement**. One possible adverse effect of volume-based measures is to encourage low-volume providers (who may also provide poorer quality of care) to increase their volume, simply to reach a threshold of 6 cases per year. Such responses would probably not improve patient outcomes to the same extent as moving patients from low-volume to high-volume hospitals. At the extreme, hospitals may loosen eligibility criteria and perform procedures on patients who are marginal or inappropriate candidates. The alternative of shutting down low-volume hospitals and transferring procedures to high-volume hospitals may overload these providers and impair access to care.

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

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Procedure volume* | | |
| Raw mean volume/standard deviation | 2 (3) | |
| Median/90th/95th percentile | 1 / 4 / 6 | |
| Stability over time, mean in 1995 / mean in 1997 | 2 / 2 | Stable |
| *Percentage of procedures at high volume hospitals* | | *Low* |
| Percentage threshold 1 (% hosp at threshold) | 40% (9%) | |
| Percentage threshold 2 (% hosp at threshold) | 34% (6%) | |
| *Persistence of high volume* | | *Low* |
| High volume remaining high, 95/96 (Threshold 1, threshold 2) | 50% / 57% | |

| | |
|---|---|
| High volume remaining high, 96/97 (Threshold 1, threshold 2) | 87% / 76% |

**Procedure Volume.** In 1996, 265 hospitals (19.7% of providers) performed at least one procedure. Of these hospitals, the mean (standard deviation) of the number of procedures was 2(3). The median was 1, and the 90$^{th}$ and 95$^{th}$ percentile was 4 and 6, respectively. In general, there are a moderate number of hospitals with lower volumes and a few hospitals with much higher volumes. Overall procedure volume was stable over the 1995-1997 time period. The mean number of procedures performed in 1995 and 1997 was 2.

**Percentage of Procedures at High Volume Hospitals.** A low percentage of procedures were performed at high volume hospitals. At the threshold 1, 39.5% esophageal resection procedures were performed at 'high volume' providers (and 8.6% of providers are 'high volume'). At the threshold 2, 34.3% were performed at 'high volume' providers (and 6.4% of providers are 'high volume').

**Persistence of High Volume.** High volume status was not persistent over time. At threshold 1, 50.0% of high volume providers in 1995 were also high volume in 1996. Similarly, 58.3% of high volume providers in 1996 were also high volume in 1997. At threshold 2, 57.1% of high volume providers in 1995 were also high volume in 1996. Similarly, 60.0% of high volume providers in 1996 were also high volume in 1997.

**Construct validity**. We estimated the correlation between esophageal resection volume and mortality, adjusting for patient characteristics such as age, sex, and APR-DRG. Volume for esophageal resection is moderately and negatively correlated with mortality for esophageal resection ($r=-.29$, $p<.05$), as well as mortality after other cancer resection procedures.

***Discussion***

Esophageal resection is a complex cancer surgery, requiring great technical skill. However, this procedure is rare, with most hospitals performing less than 10 over a 5 year period. Our empirical analyses found that the mean number of procedures per year to be 2. Relatively few hospitals actually perform this procedure, and over half only perform one per year. Despite the rarity of this procedure, relatively strong relationships between volume and outcome, specifically post-operative mortality have been noted in the literature. Nonetheless, no clear threshold has been identified. Our empirical results found volume to be moderately negatively correlated with resection mortality. However, our results do not include the complex risk adjustment used in the studies reported in the literature.

While most volume indicators are measured with high precision, the relative rarity of this procedure results in a less precise indicator, though still highly adequate for use as a quality indicator. From year to year, volumes may change, as may high volume status, as noted in our empirical analysis. Thus, if possible hospitals should examine more than one year of data, averaging volumes for a more precise estimate. Hospitals may also consider use with the pancreatic resection indicator, another complex cancer surgery.

The volume-outcome relationship on which this indicator is based may not hold over time, as providers become more experienced or as technology changes. It is important then to revisit the volume-outcome relationship to ensure the validity of this indicator. Overall, volume measures are not direct measures of quality, and are relatively insensitive. For this reason they should be used with caution and in conjunction with measures of mortality and of quality of care within that field (in this case complex cancer surgery), to ensure that increasing volumes truly improve patient outcomes.

Our empirical analyses found that relatively few resection procedures are actually performed at high volume hospitals already. However, it is unlikely that shifting the small number of procedures from low to high volume hospitals would actually increase quality of care, as frail patients would need to travel farther. Further, only a handful hospitals are high volume. It is unclear whether simply increasing volume at low volume hospitals would actually improve outcomes.

Overall, this indicator is recommended for inclusion in the HCUP II QI set. Specific caveats should be kept in mind when using this indicator. As a volume indicator, this indicator is a proxy measure for quality, and it is recommended that it be used with other indicators. Theoretically, providers could increase the number of procedures on patients with questionable indications as well, though this is more difficult than for other indicators.

**INDICATOR 5: PANCREATIC RESECTION VOLUME**

| | |
|---|---|
| Indicator | Pancreatic resection raw volume |
| Relationship to Quality | Better outcomes have been associated with higher volumes. Higher volumes thus represent better quality. |
| Benchmark | Threshold 1: 10 or more procedures per year[199]<br>Threshold 2: 11 or more procedures per year[199] |

Method:

| | |
|---|---|
| Quality Measure | Provider level pancreatic resection raw volume |
| Outcome of Interest | Discharges with ICD-9 Codes 56.2 or 52.7 in any procedure field and diagnosis code of pancreatic cancer in any field.<br><br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |
| Population at Risk | Not applicable |

***Evidence from the literature***

**Face validity**. Procedure volume is a surrogate measure of quality; its face validity depends on whether a strong association with outcomes of care is both plausible and widely accepted in the professional community.

Pancreatic cancer surgery requires technical proficiency; errors in surgical technique or management may lead to clinically significant complications, such as sepsis, anastomotic breakdown, and death. However, we are not aware of any consensus guidelines or recommendations regarding minimum procedure volume. The National Cancer Policy Board of the Institute of Medicine and the National Research Council recommends that cancer "patients undergoing procedures that are technically difficult to perform and have been associated with higher mortality in lower-volume settings (including pancreatic resection) receive care at facilities with extensive experience (e.g., high-volume facilities)."

**Precision**. The number of pancreatectomies is measured accurately with discharge data; in fact, discharge data are probably the best available source for hospital volume information. Although a few facilities have relatively high volumes, most (e.g., 263 of 298 California hospitals)[199] perform 10 or fewer esophagectomies for cancer during a 5-year period. As a result, this measure is expected to have poor precision.

**Minimum bias**. Volume measures are not subject to bias due to disease severity and comorbidities. For this reason, risk-adjustment is not appropriate. Although volume measures are theoretically subject to bias due to variation across hospitals in the use of outpatient surgery facilities, less than 1% of pediatric heart surgeries in 1996 were performed in ambulatory settings." [297]

**Construct validity**. Volume is not a direct measure of the quality or outcomes of care. Although higher volumes have been repeatedly associated with better outcomes after esophageal surgery, these findings may be limited by inadequate risk adjustment.

Only one study used clinical data to estimate the association between hospital volume and mortality following esophageal cancer surgery.[65] Begg et al. analyzed retrospective cohort

data from the Surveillance, Epidemiology, and End Results(SEER)-Medicare linked database from 1984 through 1993. The crude 30-day mortality rate was 12.9% at hospitals that performed 1-5 pancreatic resections on Medicare patients during the study period, versus 7.7% and 5.8% at hospitals that performed 6-10 and 11 or more pancreatic resections, respectively. The association between volume and mortality remained significant (p=.01) in a multivariate model, adjusting for the number of comorbidities, cancer stage and volume, and age.

Eight of the ten studies using hospital discharge data found similar effects of hospital volume. Using 1990-94 data from California, Glasgow and Mulvihill [199] estimated risk-adjusted mortality rates of 14%, 10%, 9%, 7%, 8%, and 4% across six hospital volume categories (e.g., 1-5, 6-10, 11-20, 21-30, 31-50, and >50 procedures during the 5-year study period). Their risk adjustment was quite limited; only the year of surgery, age, sex, race, payer source, extent of resection, and the total number of secondary diagnoses were included. Using 1990-97 data on radical pancreaticoduodenectomies from Maryland (adjusting only for age and payer source), Gordon et al. ([322]) estimated that the adjusted odds of death at minimal-volume (<11 "complex gastrointestinal procedures"/year) and low-volume (11-20 procedures/year) hospitals were 12.5 and 10.4 times that at a high-volume hospital (214 procedures/year). However, the generalizability of these results is limited by the fact that the last category included only one hospital.

Lieberman et al. [324] used 1984-91 hospital discharge data from New York State to analyze the association between mortality after pancreatic cancer resection and both physician and hospital volumes. Adjusting for the year of surgery, age, sex, race, payer source, transfer status, and the total number of secondary diagnoses, the standardized mortality rate was 19%, 12%, 13%, and 6% at minimal (<10 patients during the 8-year study period), low (10-50 patients), medium (51-80 patients), and high-volume (>80 patients) hospitals, respectively. Surgeon volume was less significantly associated with mortality (6-13% risk-adjusted mortality across 3 volume categories); this effect disappeared in a model that included both physician and hospital volume. The dominance of hospital volume over surgeon volume was confirmed by Sosa et al. [325], using Maryland data.

Studies using administrative data from Ontario,[326] the United Kingdom,[327] and Medicare[328] have generated results similar to those from California and New York. The only studies that failed to show a significant hospital volume-outcome association were based on relatively small, nonrepresentative samples from Department of Defense[329] or major university[330] hospitals.

Although volume-outcome associations have been demonstrated for pancreatic cancer surgery, volume seems likely to both insensitive and nonspecific as a measure of quality. It has been estimated that shifting patients in California from low-volume to high-volume hospitals would avert only 20 deaths per year, although 57% of all operations are performed in low-volume hospitals.[190] However, Gordon et al.[331] estimated that 61% of the observed reduction in statewide deaths among patients undergoing the Whipple procedure was attributable to the increasing market percentage of one facility, from 20.7% to 58.5% between 1984 and 1995.

**Fosters true quality improvement**. One possible adverse effect of volume-based measures is to encourage low-volume providers (who may also provide poorer quality of care) to increase their volume, simply to reach a threshold of 10 cases per year. Such responses would probably not improve patient outcomes to the same extent as moving patients from low-volume to high-volume hospitals. At the extreme, hospitals may loosen eligibility criteria and perform procedures on patients who are marginal or inappropriate candidates. The alternative of shutting

down low-volume hospitals and transferring procedures to high-volume hospitals may overload these providers and impair access to care.

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

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Procedure volume* | | |
| Raw mean volume/standard deviation | 3 (8) | |
| Median/90th/95th percentile | 2 / 5 / 10 | |
| Stability over time, mean in 1995 / mean in 1997 | 3 / 3 | Stable |
| *Percentage of procedures at high volume hospitals* | | *Low* |
| Percentage threshold 1 (% hosp at threshold) | 30% (5%) | |
| Percentage threshold 2 (% hosp at threshold) | 27% (4%) | |
| *Persistence of high volume* | | *Low/ Mod* |
| High volume remaining high, 95/96 (Threshold 1, threshold 2) | 73% / 74% | |
| High volume remaining high, 96/97 (Threshold 1, threshold 2) | 71% / 82% | |

**Procedure Volume.** In 1996, 429 hospitals (31.9% of providers) performed at least one procedure. Of these hospitals, the mean (standard deviation) of the number of procedures was 3 (8). The median was 2, and the 90th and 95th percentile was 5 and 10, respectively. In general, there are a moderate number of hospitals with lower volumes and a few hospitals with much higher volumes. Overall procedure volume was stable over the 1995-1997 time period. The mean number of procedures performed in 1995 and 1997 was 3.

**Percentage of Procedures at High Volume Hospitals.** A low percentage of procedures were performed at high volume hospitals. At the threshold 1, 30.3% of pancreatic resection procedures were performed at 'high volume' providers (and 5.1% of providers are 'high volume'). At the threshold 2, 27.0% were performed at 'high volume' providers (and 4.2% of providers are 'high volume').

**Persistence of High Volume.** High volume status was not persistent over time. At threshold 1, 72.7% of high volume providers in 1995 were also high volume in 1996. Similarly, 71.4% of high volume providers in 1996 were also high volume in 1997. At threshold 2, 73.7% of high volume providers in 1995 were also high volume in 1996. Similarly, 82.4% of high volume providers in 1996 were also high volume in 1997.

**Construct validity.** We estimated the correlation between pancreatic resection volume and mortality, adjusting for patient characteristics such as age, sex, and APR-DRG. Volume for esophageal resection is moderately and negatively correlated with mortality for esophageal resection ($r=-.41$, $p<.001$), as well as mortality after other cancer resection procedures.

### *Discussion*

Pancreatic resection is a complex cancer surgery, requiring great technical skill. However, this procedure is rare, with most hospitals performing less than 10 over a 5 year period. Our empirical analyses found that the mean number of procedures per year to be 3. Relatively few hospitals actually perform this procedure, and over half only perform only 2 per

year. Despite the rarity of this procedure, relatively strong relationships between volume and outcome, specifically post-operative mortality have been noted in the literature. However, no clear threshold has been identified. Our empirical analyses found pancreatic resection volume to be modestly negatively correlated with resection mortality. However, our results do not include the complex risk adjustment contained in the studies reported in the literature.

While most volume indicator are measured with high precision, the relative rarity of this procedure results in a less precise indicator, though still highly adequate for use as a quality indicator. From year to year, volumes may change, as may high volume status, as noted in our empirical analysis. Thus, if possible hospitals should examine more than one year of data, averaging volumes for a more precise estimate. Hospitals may also consider use with the esophageal resection indicator, another complex cancer surgery.

The volume-outcome relationship on which this indicator is based may not hold over time, as providers become more experienced or as technology changes. It is important then to revisit the volume-outcome relationship to ensure the validity of this indicator. Overall, volume measures are not direct measures of quality, and are relatively insensitive. For this reason they should be used with caution and in conjunction with measures of mortality and of quality of care within that field (in this case complex cancer surgery), to ensure that increasing volumes truly improve patient outcomes.

Our empirical analyses found that relatively few resection procedures are actually performed at high volume hospitals already. However, it is unlikely that shifting the small number of procedures from low to high volume hospitals would actually increase quality of care, as frail patients would need to travel farther. Further, only a handful hospitals are high volume. It is unclear whether simply increasing volume at low volume hospitals would actually improve outcomes.

Overall, this indicator is recommended for inclusion in the HCUP II QI set. Specific caveats should be kept in mind when using this indicator. As a volume indicator, this indicator is a proxy measure for quality, and it is recommended that it be used with other indicators. Theoretically, providers could increase the number of procedures on patients with questionable indications as well, though this is more difficult than for other indicators.

**INDICATOR 6: PEDIATRIC HEART SURGERY VOLUME**

| | |
|---|---|
| Indicator | Pediatric heart surgery raw volume |
| Relationship to Quality | Better outcomes have been associated with higher volumes. Higher volumes thus represent better quality. |
| Benchmark | Threshold : 100 or more procedures per year[194] |

Method:

| Quality Measure | Provider level pediatric heart surgery raw volume |
|---|---|
| Outcome of Interest | Discharges with ICD-9 procedure codes for 1) specified heart surgery (see Appendix 6) in any field or 2) procedure code for any heart surgery and diagnosis of hypoplastic left heart syndrome in any field.<br><br>Age less than 18.<br><br>See appendix for additional exclusions.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |
| Population at Risk | Not applicable |

***Evidence from the literature***

**Face validity**. Procedure volume is a surrogate measure of quality; its face validity depends on whether a strong association with outcomes of care is both plausible and widely accepted in the professional community.

Pediatric cardiac surgery requires technical proficiency with the use of complex equipment. Technical errors may lead to clinically significant complications, such as arrhythmias, congestive heart failure, and death. However, we are not aware of any consensus guidelines or recommendations regarding minimum procedure volume.

**Precision**. The number of pediatric cardiac procedures is measured accurately with discharge data; in fact, discharge data are probably the best available source for hospital volume information. Previous studies suggest that pediatric cardiac surgery is already highly concentrated at a relatively small number of facilities (e.g., 16 hospitals in New York, 37 in California and Massachusetts together). Although some of these facilities have very high volumes, a significant number (e.g., 16 hospitals in California and Massachusetts) perform fewer than 10 cases per year. The highly skewed volume distribution may have an adverse effect on the precision of this measure.

**Minimum bias**. Volume measures are not subject to bias due to disease severity and comorbidities. For this reason, risk-adjustment is not appropriate. Less than 1% of pediatric heart surgery are performed on an outpatient basis. [297]

**Construct validity**. Volume is not a direct measure of the quality or outcomes of care. Although higher volumes have been repeatedly associated with better outcomes after pediatric cardiac surgery, these findings may be limited by inadequate risk adjustment.

Only one study used prospectively collected clinical data to estimate the association between hospital volume and mortality following pediatric cardiac surgery.[194] Hannan et al. ordered all cardiac surgical procedures by their actual mortality rates in the 1992-95 Cardiac Surgery Reporting System database. Expert clinicians then grouped the procedures into four clinically sensible subgroups, designed to achieve maximal separation of crude mortality rates (from 1.4% for Category I to 20.1% for Category IV). A multivariate model that included age, complexity category, and four comorbidities (preoperative cyanosis or hypoxia, acidemia, pulmonary hypertension, major extracardiac anomalies) achieved excellent calibration and discrimination (c=0.818). Using this model to estimate risk-adjusted mortality, Hannan et al. found a statistically significant hospital effect (8.26% risk-adjusted mortality at hospitals with fewer than 100 cases per year, versus 5.95% at higher volume hospitals), which was limited to surgeons who performed at least 75 cases per year. Lower volume surgeons experienced relatively high mortality, regardless of total hospital volume. Risk-adjusted mortality differed between low and high-volume hospitals for all 4 complexity categories, although the smallest difference occurred for the highest risk procedures.

Two other studies using hospital discharge data found similar effects of hospital volume. Using aggregated data from California (1988) and Massachusetts (1989), Jenkins et al.[332] estimated risk-adjusted mortality rates of 8.35% and 5.95% at low-volume (100 or fewer cases) and high-volume (more than 100 cases), respectively. However, they also demonstrated especially high risk-adjusted mortality (18.5%) at very low-volume hospitals with fewer than 10 annual cases, and especially low mortality (3.0%) at very high-volume hospitals with more than 300 annual cases. Jenkins et al. could not evaluate the impact of surgeon volume, but they did report stronger volume effects for higher-risk procedures (e.g., OR=12.1 and 3.2 for Category III-IV procedures at hospitals with <10 and 10-100 annual cases, versus OR=2.4 for Category I-II procedures at hospitals with 10-100 annual cases). Finally, Sollano et al. [295] applied the same 4-category risk adjustment procedure developed by Jenkins to hospital discharge data from New York State in 1990-95. They reported a modest but statistically significant effect (OR=0.944 for each additional 100 annual cases), which was limited to neonates (OR=0.636) and post-neonatal infants (OR=0.720) in stratified analyses.

Although volume-outcome associations have been demonstrated for pediatric cardiac surgery, volume seems likely to both insensitive and nonspecific as a measure of quality. In addition, pediatric cardiac care is already regionalized, so most procedures are performed in medium-to-high volume hospitals. It has been estimated that shifting patients in California from low-volume to high-volume hospitals would avert only 7 deaths per year.[190]

**Fosters true quality improvement**. One possible adverse effect of volume-based measures is to encourage low-volume providers (who may also provide poorer quality of care) to increase their volume, simply to reach a threshold of 100 cases per year. Such responses would probably not improve patient outcomes to the same extent as moving patients from low-volume to high-volume hospitals. At the extreme, hospitals may loosen eligibility criteria and perform procedures on patients who are marginal or inappropriate candidates. The alternative of shutting down low-volume hospitals and transferring procedures to high-volume hospitals may overload these providers and impair access to care.

**Prior use.** Pediatric cardiac surgical volume has not been widely used as an indicator of quality.

***Empirical Evidence***

| Test | Statistic | Rating |
|---|---|---|
| *Procedure volume* | | |
| Raw mean volume/standard deviation | 53 (90) | |
| Median/90th/95th percentile | 2.5 / 149 / 245 | |
| Stability over time, mean in 1995 / mean in 1997 | 52 / 52 | Stable |
| *Percentage of procedures at high volume hospitals* | | *Moderate* |
| Percentage threshold 1 (% hosp at threshold) | 76% (21%) | |
| Percentage threshold 2 (% hosp at threshold) | N/A | |
| *Persistence of high volume* | | *Moderate* |
| High volume remaining high, 95/96 (Threshold 1, threshold 2) | 85% | |
| High volume remaining high, 96/97 (Threshold 1, threshold 2) | 84% | |

**Procedure Volume.** In 1996, 126 hospitals (9.3% of providers) performed at least one procedure. Of these hospitals, the mean (standard deviation) of the number of procedures was 53 (90). The median was 2.5, and the 90th and 95th percentile was 149 and 245, respectively. In general, there are many hospitals with lower volumes and a few hospitals with much higher volumes. Overall procedure volume was stable over the 1995-1997 time period. The mean number of procedures performed in 1995 and 1997 was 52.

**Percentage of Procedures at High Volume Hospitals.** A moderate percentage of procedures were performed at high volume hospitals. At the threshold 1, 75.5% of pediatric heart surgeries were performed at 'high volume' providers (and 21% of providers are 'high volume'). There is no threshold 2 for this procedure.

**Persistence of High Volume**. High volume status was moderately persistent over time. At threshold 1, 84.6% of high volume providers in 1995 were also high volume in 1996. Similarly, 84.0% of high volume providers in 1996 were also high volume in 1997.

**Construct validity.** We estimated the correlation between pediatric heart surgery volume and mortality, adjusting for patient characteristics such as age, sex, and APR-DRG. Pediatric heart surgery volume is independently and negatively correlated with mortality ($r=-.27$, $p<.05$). However, this analysis does not include the intensive risk adjustment included in the volume studies described in the literature review.

### *Discussion*

Pediatric Heart surgery includes a number of procedures, varying in difficulty. In general pediatric heart surgery is technically complex and differs from adult heart surgery. A large number of hospitals perform at least one procedure, but only 21% of hospitals are actually high volume. Further, over half of provider perform only 2.5 procedures a year. Higher volumes of pediatric heart surgery have been associated with better outcomes, namely fewer deaths. Providers may want to examine operator volume as well as hospital volume. Our empirical results noted that pediatric heart surgery volume was slightly positively correlated with mortality. Note that the correlation, significant due to the large number of observations, is small and not considered important. Our results do not include the complex risk adjustment contained in the studies reported in the literature, and required for this comparison.

This indicator is measured with great precision, as is expected with all volume indicators. It is expected that volume for pediatric heart surgery would be measured precisely using discharge abstract data. Most procedures are performed in an inpatient setting.

The volume-outcome relationship on which this indicator is based may not hold over time, as providers become more experienced or as technology changes. It is important then to revisit the volume-outcome relationship to ensure the validity of this indicator. Overall, volume measures are not direct measures of quality, and are relatively insensitive. For this reason they should be used with caution and in conjunction with measures of mortality and of quality of care within that field (in this case pediatric surgery), to ensure that increasing volumes truly improve patient outcome.

Our empirical analyses found that about ¾ of pediatric heart surgeries are actually performed at high volume hospitals already, suggesting regionalization. This leaves little room for improvement. It is unclear whether simply increasing volume at low volume hospitals would actually improve outcomes. It is possible that hospitals could increase volume simply by increasing the number of borderline or inappropriate procedures performed. For this reason, it may be prudent to examine this indicator alongside area rates for this procedure, and examinations of the appropriateness of the procedures.

Overall, this indicator is recommended for inclusion in the HCUP II QI set. Specific caveats should be kept in mind when using this indicator. As a volume indicator, this indicator is a proxy measure for quality, and it is recommended that it be used with other indicators. Theoretically, providers could increase the number of procedures on patients with questionable indications as well, though this is more difficult than for other indicators.

**INDICATOR 7: PERCUTANEOUS TRANSLUMINAL CORONARY ANGIOPLASTY (PTCA) VOLUME**

| | |
|---|---|
| Indicator | Percutaneous transluminal coronary angioplasty (PTCA) raw volume |
| Relationship to Quality | Better outcomes have been associated with higher volumes. Higher volumes thus represent better quality. |
| Benchmark | Threshold 1: 200 or more procedures per year[191]<br>Threshold 2: 400 or more procedures per year[61, 192] |

Method:

| | |
|---|---|
| Quality Measure | Provider level PTCA raw volume |
| Outcome of Interest | Discharges with ICD-9 Codes 36.01, 36.02, 36.05 or 36.06 in any procedure field.<br><br>Age 40 years and older.<br><br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |
| Population at Risk | Not applicable |

***Evidence from the literature***

**Face validity**. Procedure volume is a surrogate measure of quality; its face validity depends on whether a strong association with outcomes of care is both plausible and widely accepted in the professional community.

PTCA is a procedure that requires technical proficiency with the use of complex equipment. Technical errors may lead to clinically significant complications, such as abrupt coronary occlusion with or without myocardial infarction, emergency coronary bypass surgery, and death. On the basis of this knowledge and empirical literature (summarized below), the American Heart Association (AHA) and the American College of Cardiology (ACC) have stated that "a significant number of cases per institution – at least 200 PTCA procedures annually – is essential for the maintenance of quality and safe care."[191] More recent literature (summarized below) led a subsequent expert panel to recommend that "an institution should have an activity level of at least 400 coronary procedures/year…an institution with a volume of <200 procedures/year, unless in a region that is underserved because of geography, should carefully consider whether it should continue to offer the service." [333]

The same task force expressed concern that "a majority of operators fail to meet the requirements for maintenance of competence, which is a minimum of 75 PTCA procedures performed per year as the primary operator." This standard has been endorsed by the American College of Physicians;[334] the Society for Cardiac Angiography proposed a lower minimum of 50 cases per year. [334]

**Precision**. PTCA is an increasingly common procedure; approximately 452,000 were performed in the USA in 1997 (16.7 per 10,000 persons).[316] In the Nationwide Inpatient Sample from the Healthcare Cost and Utilization Project, 214 hospitals reported a mean of 382 PTCAs per year in 1993-94. The 27% of hospitals that were classified as low-volume (<200 per year) performed 5% of the procedures, whereas the 31% of hospitals classified as medium-volume

(201-400 per year) performed 21% of the procedures and the 42% of hospitals classified as high-volume (>400 per year) performed 74% of the procedures.[172] Based on state all-payer databases, the mean annual frequency of angioplasties was 226 per hospital in California in 1989[335] and 505 per hospital in New York in 1991-1994.[61]

The number of PTCA procedures is measured accurately with discharge data; in fact, discharge data are probably the best available source for hospital volume information. The large number of procedures performed annually at most hospitals suggests that annual volume is not subject to considerable random variation.

**Minimum bias**. Volume measures are not subject to bias due to disease severity and comorbidities. For this reason, risk-adjustment is not appropriate. Although volume measures are theoretically subject to bias due to variation across hospitals in the use of outpatient surgery facilities, only 7.6% of PTCAs in 1996 were performed in ambulatory settings.[297]

**Construct validity**. Volume is not a direct measure of the quality or outcomes of care. Although higher volumes have been repeatedly associated with better outcomes after PTCA, these findings may be limited by inadequate risk adjustment. Using hospital discharge data to adjust for age, gender, multivessel angioplasty, unstable angina, and 6 comorbidities, high-volume hospitals had significantly lower rates of same-stay coronary bypass surgery (CABG) and inpatient mortality than low-volume hospitals.[172] Although the magnitudes of the adjusted differences were not reported, the unadjusted differences were modest (e.g., 3.8% versus 4.6% mortality and 4.3% versus 4.6% CABG rates after myocardial infarction, 0.8% versus 1.0% mortality and 2.8% versus 4.0% CABG rates without myocardial infarction). An earlier study using similar data and volume thresholds reported more adverse outcomes (e.g., CABG or death) than expected at low-volume hospitals (e.g., 12.4% observed versus 10.0% expected after myocardial infarction, 6.3% observed versus 5.2% expected without myocardial infarction) and fewer adverse outcomes than expected at high-volume hospitals (e.g., 8.3% observed versus 10.5% expected after myocardial infarction, 4.4% observed versus 5.0% expected without myocardial infarction).[235] A study based on Medicare data also reported a significant association between hospital volume and mortality, after adjustment for age, sex, race, and year, although no adjusted measures of effect were reported.[112, 172, 192, 237, 336, 337] Better studies based on clinical data systems (adjusting for left ventricular function) have confirmed higher risk-adjusted mortality and CABG rates at low and medium (<400 cases per year) volume hospitals, relative to high-volume hospitals (e.g., 1.1% versus 0.80-0.95% mortality, 4.2% versus 2.8-3.7% CABG).[61] A similar study based on clinical data from the Society for Cardiac Angiography and Interventions confirmed the validity of a higher volume threshold than the 200 cases per year originally recommended by the AHA and ACC. Adjusted odds ratios for post-PTCA complications (e.g., death, emergency CABG, or myocardial infarction) were 1.14 at hospitals with 200-399 cases per year, 0.66 at hospitals with 400-599 cases, and 0.54 at hospitals with 600 or more cases.[192]

Studies of operator volume are less directly relevant to the HCUP Quality Indicator project, but three studies (from New York, Northern New England, and a community hospital in Los Angeles) have supported associations between operator volume and angiographic and clinical success rates,[336] as well as risk-adjusted same-stay CABG rates.[61, 64, 258, 310-312] These studies did not demonstrate any association between operator volume and inpatient mortality. Finally, the most recent study showed that the associations between operator volume and both clinical success rates and CABG rates apparently disappeared between 1990-1993 and 1994-1996 in northern New England.[338]

Although volume-outcome associations have been demonstrated for PTCA, volume seems likely to both insensitive and nonspecific as a measure of quality. For example, Hannan demonstrated that low-volume cardiologists can achieve excellent outcomes at high-volume hospitals (e.g., 0.7% risk-adjusted mortality, 2.9% same-stay CABG).[61] Nonetheless, it has been estimated that shifting patients in California from low-volume to high-volume hospitals would avert 80 deaths per year. This number is consistent with a national estimate of 137 averted deaths and 404 averted same-stay CABG.[172] It is possible that a low-volume provider may be unavoidable for urgent procedures in less populated areas. But it is unclear whether such urgent cases would do better with (low-volume) PTCA than with alternative, non-PTCA treatments that are not as volume-sensitive.

**Fosters true quality improvement**. One possible adverse effect of volume-based measures is to encourage low-volume providers (who may also provide poorer quality of care) to increase their volume, simply to reach a threshold of 200 or 400 cases per year. Such responses would probably not improve patient outcomes to the same extent as moving patients from low-volume to high-volume hospitals. At the extreme, hospitals may loosen eligibility criteria and perform procedures in patients who are marginal or inappropriate candidates. The alternative of shutting down low-volume hospitals and transferring procedures to high-volume hospitals may overload these providers and impair access to care.

**Prior use.** PTCA volume has not been widely used as an indicator of quality, although specific volume thresholds have been suggested as "standards" for the profession.[333] In its Web site, the Pacific Business Group on Health[304] (http://www.healthscope.org) states that "one marker of how well a hospital is likely to perform is the experience of the hospital and its surgical team...in the absence of data to compare hospitals on their complications and survival rates, you can begin evaluating experience by looking at the number of (PTCA) surgeries a hospital performs each year." The Center for Medical Consumers posts hospital-specific and operator-specific PTCA volumes for New York hospitals (http://www.medicalconsumers.org).

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Procedure volume* | | |
| Raw mean volume/standard deviation | 418 (400) | |
| Median/90th/95th percentile | 330 / 869 / 1157 | |
| Stability over time, mean in 1995 / mean in 1997 | 379 / 447 | Growing |
| *Percentage of procedures at high volume hospitals* | | *High/ Mod* |
| Percentage threshold 1 (% hosp at threshold) | 96% (69%) | |
| Percentage threshold 2 (% hosp at threshold) | 77% (42%) | |
| *Persistence of high volume* | | *High* |
| High volume remaining high, 95/96 (Threshold 1, threshold 2) | 98% / 98% | |
| High volume remaining high, 96/97 (Threshold 1, threshold 2) | 98% / 97% | |

**Procedure Volume**. In 1996, 365 hospitals (26.4% of providers) performed at least one procedure. Of these hospitals, the mean (standard deviation) of the number of procedures was 418 (400). The median was 330, and the 90th and 95th percentile was 869 and 1157, respectively. In general, there are a moderate number hospitals with lower volumes and a few hospitals with

much higher volumes. Overall procedure volume grew over the 1995-1997 time period. The mean number of procedures performed in 1995 and 1997 was 379 and 447, respectively.

**Percentage of Procedures at High Volume Hospitals.** A moderate to high percentage of procedures were performed at high volume hospitals, depending on the threshold. At the threshold 1, 95.7% of PTCA procedures were performed at 'high volume' providers (and 69% of providers are 'high volume'). At the threshold 2, 77.0% were performed at 'high volume' providers (and 42% of providers are 'high volume').

**Persistence of High Volume**. High volume status was highly persistent over time. At threshold 1, 97.8% of high volume providers in 1995 were also high volume in 1996. Similarly, 97.9% of high volume providers in 1996 were also high volume in 1997. At threshold 2, 98.5% of high volume providers in 1995 were also high volume in 1996. Similarly, 96.5% of high volume providers in 1996 were also high volume in 1997.

**Construct validity**. As we did not retain the PTCA mortality indicator due to inadequate precision, we were unable to test the construct validity of this indicator. However, PTCA volume is negatively related to several other post-procedural mortality rates: CABG ($r=-.21$, $p<.001$), craniotomy ($r=-.200$, $p<.0001$), and AAA repair ($r=-.45$, $p<.0001$).

***Discussion***

Percutaneous transluminal coronary angioplasty (PTCA) is a relatively common procedure. Our empirical analysis found a mean of 418 procedures per year. A substantial number of hospitals perform at least one procedure, but only 69% (threshold 1) or 42% (threshold 2) of hospitals are actually high volume. Higher volumes of PTCA have been associated with better outcomes, namely fewer deaths and post-procedural coronary artery bypass grafts (CABG). The AHA/ACC have suggested that hospitals perform at least 200 PTCA procedures per year to maintain proficiency. Though many hospitals meet and exceed this rate, there are still hospitals that do not meet this guideline. Operator volume is also important, since many operators do not meet the 75 procedure minimum suggested. Providers may wish to examine operator rates alongside this indicator.

This indicator is measured with great precision, as is expected with all volume indicators. It is expected that volume for PTCA would be measured precisely using discharge abstract data. Though most procedures are performed on an inpatient basis, about 7% of procedures are performed on an outpatient basis. Providers may wish to examine outpatient and inpatient rates together.

The volume-outcome relationship on which this indicator is based may not hold over time, as providers become more experienced or as technology changes. It is important then to revisit the volume-outcome relationship to ensure the validity of this indicator. Overall, volume measures are not direct measures of quality, and are relatively insensitive. For this reason they should be used with caution and in conjunction with measures of mortality and of quality of care within that field (in this case cardiac care), to ensure that increasing volumes truly improve patient outcomes.

Our empirical analyses found that most PTCA procedures are actually performed at high volume hospitals already. This leaves little room for improvement. Further, many hospitals are high volume. It is unclear whether simply increasing volume at low volume hospitals would actually improve outcomes. It is possible that hospitals could increase volume simply by increasing the number of borderline or inappropriate procedures performed. For this reason, it

may be prudent to examine this indicator alongside area rates for this procedure, and examinations of the appropriateness of the procedures.

Overall, this indicator is recommended for inclusion in the HCUP II QI set. Specific caveats should be kept in mind when using this indicator. As a volume indicator, this indicator is a proxy measure for quality, and it is recommended that it be used with other indicators. Providers could increase the number of procedures on patients with questionable indications as well, without improving quality of care.

## Process Measures

## 3.E.2 Provider-Level Utilization Measures

**INDICATOR 8: CESAREAN SECTION DELIVERY RATE**

| | |
|---|---|
| Indicator | Cesarean section delivery rate |
| Relationship to Quality | C-section has been identified as an overused procedure. As such, lower rates of cesarean section represent better quality. |
| Benchmark | State, regional or peer group average<br>HP 2010 goal: 15 c-sections per 100 births. |

Method:

| | |
|---|---|
| Quality Measure | Provider-level number of C-sections per 100 deliveries (see Appendix 6). |
| Outcome of Interest | Number of C-sections. |
| Population at Risk | All deliveries (see Appendix 6). |

***Evidence from the literature***

**Face validity.** The rate of cesarean delivery in the United States increased from 5.5% in 1970 to a high of 24.7% in 1988, with a subsequent decrease to 20.7% in 1996.[339] Demographic changes in the childbearing population likely account for a relatively small part of this increase, as suggested by Parrish et al.[340] in a study of primary cesarean delivery rates in Washington State from 1987 to 1990. Previous data in population studies have failed to document commensurate improvements in outcomes associated with this increased utilization,[341] which has raised questions regarding the appropriateness of current practices. Moreover, cesarean delivery is the most common operative procedure performed in the United States[342] and is associated with higher costs than vaginal delivery. The U.S. Department of Health and Human Services[5] and private groups[343] are encouraging reductions in cesarean rate, the former having set a goal of reducing the cesarean delivery rate to 15% by the year 2000.

While appropriateness of the procedure depends largely on patients' clinical characteristics (see Minimum bias below), studies have shown that individual physician practice patterns account for a significant portion of the variation in cesarean delivery rates.[344-349] Non-clinical factors such as patient insurance status, hospital characteristics, and geographic region have also been related to rates.[350-356]

**Precision.** Burns et al.[347] have shown cesarean delivery is common enough to make good statistical comparisons of hospital and even physician style feasible. Furthermore, the eligible population (pregnant women) is well defined and hospital level reporting reduces the small *n* problem that may occur with individual providers.

**Minimum bias.** The overall CS rate cannot determine appropriate use, but the variation in rates across institutions/regions may if the variations do not merely reflect variations in patient disease severity and co-morbidities. Comparison of measures of utilization or outcomes, to be fair, requires adequate adjustment for case mix.[8] Studies that have risk-adjusted cesarean

delivery rates differ in the risk factors included (indications for cesarean delivery surrounding many risk factors is controversial) and data sources used.

Keller et al.[357] examined singleton births greater than 2500 grams in hospitals in Washington State in 1989 and 1990 using a combination of administrative and birth certificate data. The authors developed separate multivariate models for each of 4 groups (prior cesarean, breech, first birth, other). Risk adjustment including minimal clinical detail (parity, prior cesarean, breech/malposition, placenta/cord problems, active herpes, mother's age, amnionitis, birth weight, sex of child) could explain most of variance among hospitals. The authors concluded that "adjustment of rates did not greatly alter hospital rankings, but the adjustments are fair, improve face validity, and work surprisingly well in explaining which mothers get cesareans. So they should improve the acceptance of monitoring of rates."

In contrast, Aron et al.[358] used data from standardized reviews of medical records to adjust for clinical risk factors in women without prior cesarean section who delivered in the Cleveland area from 1993 to mid 1995. With respect to unadjusted rates, 7 hospitals were statistical outliers. After risk-adjustment, outlier status changed for 5 (24%) of the 21 hospitals. When hospitals were rank-ordered on the basis of cesarean delivery rates, the correlation between unadjusted and adjusted ranking was only moderate. Hospital rankings often changed and, in 12 of the 21 hospitals (57%), the relative difference in unadjusted and adjusted rates was greater than 10%. The authors note that some of the variation in prevalence of risk factors across hospitals may reflect differences in documentation.

Similarly, Bailit et al.[359] showed that risk-adjusting primary cesarean delivery rates using a state birth certificate database substantially changes how hospital performance is judged. Specifically, they found 27% of hospitals with unadjusted rates in the top quartile had adjusted rates that were "risk-appropriate" and that 23% of hospitals with unadjusted rates that were not in the top quartile had adjusted rates that were. In another study using birth certificate data, Glantz[360] also found that crude rates do not accurately reflect the differences in cesarean delivery rates among hospitals and commented "to make judgements regarding clinical practices on the basis of unadjusted rates entails the risk of unwarranted emulation of some hospitals that only appear to have low rates and unfair criticism of some hospitals with seemingly high rates."

Gregory et al[361] used discharge data to adjust for clinical risk factors for cesarean delivery in 92,798 singleton Medicaid deliveries in Los Angeles County in 1991. The authors categorized patients according to a revised hierarchical set of indications for cesarean section,[362] revised and validated by [363] and adjusted for clinical risk factors based on the presence of maternal ICD-9-CM diagnostic codes. The aim of the LA County study was to describe the difference in risk-adjusted cesarean delivery rates according to hospital type; the effect of risk-adjustment on individual hospital rates or on hospital rankings was not evaluated.

**Construct validity.** We found no studies explicitly evaluating the construct validity of this indicator. In other words, there is no evidence that hospitals with lower cesarean rates more frequently provide better quality of care according to other measures.

As the cesarean rate for "optimal" quality care is unknown, many studies are careful to note that lower cesarean rates do not necessarily reflect better quality care. In some instances, a higher cesarean rate could reflect more appropriate use of the procedure. For example, a meta-analysis by Gifford et al.[364] suggests that elective cesarean delivery for breech presentation may be associated with better neonatal outcomes. Cesarean delivery rates substantially less than the national trend of 90% for infants in a breech presentation may indicate underutilization; however, correlation with maternal and neonatal outcomes would help clarify this issue.

**Fosters true quality improvement.** The cesarean delivery rate can be decreased by decreasing the primary cesarean delivery rate and/or increasing the vaginal birth after cesarean (VBAC) rate. In some hospitals, one or both of these might result in more maternal and/or infant complications. Sachs et al.[365] note that when a trial of labor after cesarean delivery fails, the rate of maternal morbidity, including infection and operative injuries, increases substantially. The authors cite increasing incidence of the major risk of a trial of labor, uterine rupture, in several states in recent years. However, they caution that, in the absence of chart reviews, one cannot be sure that all these cases involved rupture of a uterine scar from a previous cesarean delivery. Sachs and colleagues also express concern that attempts to decrease the primary cesarean delivery rate may lead to complications associated with higher rates of instrumented (forceps or vacuum-assisted) vaginal delivery. Studies that have compared rates of instrumented vaginal delivery by physicians with low cesarean rates to physicians with higher cesarean rates have found either no significant difference, [346, 366] or that physicians with low cesarean rates actually use instrumented delivery less.[367] Depending on how rates are risk adjusted, there is the possibility that providers will respond by upcoding diagnoses once measurement begins, though we found no studies where this has been documented. Errors leading to bias are also possible at the data abstraction level.

**Prior use.** Cesarean section rates were one of the first measures used to judge hospital and health plan performance,[368] and has become one of the most commonly used indicators. Public Citizen's Health Research Group report cesarean delivery rates for 3159 hospitals in 41 states. Cesarean delivery rate is included among the 16 core performance measures in the Maryland Hospital Association's Quality Indicator (QI) Project.[369] Repeat and all cesarean section rates are used by the University Hospital Consortium.[370] Total cesarean delivery rates are used by Florida Agency for Health Care Administration,[371] Greater New York Hospital Association,[372] Michigan Hospital Association,[373] Pacific Business Group on Health,[304] United Health Care, Cleveland Health Quality Choice,[374] JCAHO's IMSystem, Virginia Health Information,[375] Washington State Community Health Information Partnership,[376] HealthGrades.com.[377] In addition, Cesarean section was included in the previous version of HCUP I QIs, and the reduction of cesarean section rate is a goal for Healthy People 2010.[5]

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw provider level rate/standard deviation | 21.4%, 8.7% | |
| Systematic provider-level standard deviation* | 4.5% | High |
| Provider variation as a percentage of total variation* | 1.2% | High |
| Signal ratio | 88.2% | High |
| R-Square* | 92.5% | Very High |
| * age adjusted | | |
| *Minimum Bias – age risk adjustment* | | |
| Signal variance change with risk adjustment | No change | Good |
| Absolute impact: | | |
| Average absolute change (in %) | 5.4% | Very Good |
| Relative impact: | | |
| Rank correlation | 0.956 | Very Good |
| Percent remaining in high decile/low decile | 72.4% / 79.3% | Good/ Fair |
| Percent changing more than 2 deciles | 2.6% | Very Good |

**Precision.** This indicator is precise, with a raw provider level mean of 21.4% and a substantial standard deviation of 8.7%. The systematic provider level standard deviation is high, at 4.5%. The provider level variation accounts for a high of total variation, at 1.2%. This means that relative to other indicators, a higher percentage of the variation occurs at the provider level, rather than the discharge level, although more of the variation occurs at the discharge level than for some indicators. Finally, the signal ratio is high, at 88.2%. This means that it is likely that the observed differences in provider performance represent true differences in provider performance, although some of the observed differences are due to unobserved differences in patient characteristics. The R-Square for this indicator is very high at 92.5%, meaning that some additional signal can be extracted using multivariate techniques.

**Bias.** We did not perform APR-DRG risk adjustment because the categories correspond to the outcome of interest. As a result, the indicator performs well on the multiple measures of minimum bias, using age only (of the mother) risk adjustment. The rank correlation is good at 0.956. Age risk adjustment does seems to impact the lowest and highest decile, with 72.4% of providers remaining after risk adjustment in the highest decile and 79.3% remaining in the lowest decile. There does not seem to be disproportionate impact at either extreme, though the performance in the lowest decile was poorer than that of other indicators. The absolute impact was minimal.

**Construct validity.** C-section rate loads very highly on factor three, and is inversely related to vaginal delivery after cesarean section and positively related, but to a lesser extent, to incidental appendectomy.

***Discussion***

Cesarean section has been targeted as a potentially overused procedure, as the rate of c-sections has increased over the past few decades. Despite a recent decrease, many organizations have aimed to monitor and reduce the c-section rate.

Our empirical analyses demonstrated that c-section rate is measured with good precision, as would be expected from the relatively high rates of this procedure. Given the high signal ratio, it is likely that the observed differences in the provider performance represent true differences in provider performance, rather than random variation. This cannot fully account for systematic bias in the indicator, however.

While cesarean section performed well on our tests of minimum bias, with risk adjustment having only a modest impact, the only adjustment performed was of age of the mother (since there is no variation in severity within the cesarean section APR-DRG). We were unable to link maternal and infant records, which may add beneficial risk adjustment factors. Other clinical measures not available in our data set may vary systematically by hospital and introduce some bias. Our literature review located several studies of risk adjustment for this indicator, finding that risk adjustment did affect the outlier status and rankings of as many as 25% of the hospitals. Risk adjustment was also advocated as a means to make the indicator more palatable to providers. Given these results, providers may want to examine the clinical characteristics of their populations when interpreting the results of this indicator. Clinical characteristics such as prior cesarean, parity, breech presentation, placental or cord complications, STDs, infections, birth weight have been shown to explain substantial amounts of variations in cesarean section rates. Information regarding some of these factors may be available by linking maternal discharge records to birth records.

We located no additional studies that examined the construct validity of c-section. However, we found that c-section and VBAC are strongly negatively correlated, as one would expect. Increasing the VBAC rate was proposed as one means of reducing c-section rate. However, a simple correlation does not imply causation, nor guarantee that this relationship is evidence of an underlying *quality* relationship.

While Healthy People 2010 has established a goal of 15 cesarean sections per 100 births and 63 repeat cesarean sections per 100 births with previous c-section, the ideal rate of cesarean section has not been established. Providers should compare rates to other standards such as regional or national averages. High rates may be explainable by more complex case mix; as there are some cases where a trial of labor would be contraindicated. Thus providers with high c-section rates should examine the patient characteristics available in more clinically detailed records such as birth records, to establish the appropriateness of the procedures. Providers may also wish to further break down the cesarean section indicator to primary and repeat cesarean section rates.

Overall, this indicator is recommended for inclusion in the HCUP II QI set. It received an empirical rating of 17 out of 26. This indicator is recommended with several caveats of use. Potential additional bias may result from clinical differences not identifiable in administrative data, so supplemental risk adjustment with linked birth records or other clinical data may be desirable. As a utilization indicator, the construct validity relies on the actual inappropriate use of procedures in hospitals with high rates, and this should be investigated further. Finally, this caution should be maintained for cesarean rates that are drastically below or above the average or recommended rates.

**INDICATOR 9: INCIDENTAL APPENDECTOMY AMONG THE ELDERLY RATE**

| | |
|---|---|
| Indicator | Incidental appendectomy among the elderly rate |
| Relationship to Quality | Incidental appendectomy among the elderly is contraindicated. As such lower rates represent better quality care. |
| Benchmark | State, regional, peer groups average. |

Method:

| | |
|---|---|
| Quality Measure | Number of incidental appendectomies per 100 elderly with intra-abdominal procedure. |
| Outcome of Interest | Number of incidental appendectomies (see Appendix 6). |
| Population at Risk | All non-maternal/non-neonatal discharges age 65 years or older with intra-abdominal procedure in any procedure field (see Appendix 6).<br><br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates |

***Evidence from the literature***

**Face validity.** The removal of the appendix incidental to other abdominal surgery, such as urological, gynelogical, or gastrointestinal surgeries, is intended to eliminate the risk of future appendicitis, and to simplify any future differential diagnoses of abdominal pain. Controversy about the procedure has abounded since the early 20th century.[378] Evidence remains unclear for the population as a whole, whether the removal of the appendix increases risk of morbidity and mortality significantly, or whether, given the low risk for future appendicitis, and ease of treatment, it is worth any amount of extra risk. Traditionally, it has been noted that the removal of the appendix may potentially contaminate the other- wise clean operating field.

Unfortunately, the only three published randomized controlled trials did not have sufficient power to exclude a clinically meaningful adverse effect of incidental appendectomy.[379-381] In a prospective, randomized placebo-controlled trial of prophylactic antibiotics, patients who underwent cholecystectomy with incidental appendectomy had a substantially higher wound infection rate than patients who underwent cholecystectomy alone if they did not receive antibiotics (40% versus 16%), but not if they did (9% versus 10%).[382] Two retrospective studies based on large administrative data sets demonstrated significant risk associated with incidental appendectomy among cholecystectomy patients, after adjusting for age, sex, primary diagnosis, and comorbidities. The risk of wound infection was 83% higher among elderly Medicare beneficiaries,[383] and the risk of any postoperative complication was 53% higher in Ontario general hospitals,[384] when incidental appendectomy was performed. These studies demonstrated substantial selection bias, in that patients selected for incidental appendectomy were younger and had less comorbidity than other cholecystectomy patients. The difference in baseline characteristics was so profound that incidental appendectomy was associated with a 63% decrease in the unadjusted risk of death in Ontario; this effect disappeared after risk-adjustment and reversed after stratification. These results raise serious questions about

the validity of other nonrandomized studies of incidental appendectomy, which used smaller samples and inadequate risk adjustment.

Many of the studies discussed above group all patients together regardless of age. However, Andrew and Roty [385] showed that incidental appendectomy was associated with a higher risk of wound infection (5.9% versus 0.9%) among cholecystectomy patients who were at least 50 years of age, but not among younger patients. Based on this finding, and the findings of Warren and colleagues, most commentators believe that incidental appendectomy is inappropriate for elderly patients.[386-388] Although elderly individuals have a higher risk of serious complications due to a perforated appendix, the probability of developing appendicitis is very low, with a lifetime risk of less than 1%. In this age group, it would require at least 115 incidental appendectomies to prevent one hospitalization for appendicitis, and 4,472 incidental appendectomies to avoid a single death.[383, 389] Given this logic, the risk of incidental appendectomy is believed to outweigh the benefits in the elderly population.

**Precision.** Rates of incidental appendectomy in the elderly have not been widely studied recently. One 1993 study of Medicare beneficiaries found that about 4% of cholecystectomy cases had a secondary procedure code of incidental appendectomy. [383] Another study of leading urological surgery departments (non-random sample) found that over 50% of departments did not routinely perform incidental appendectomy during radical cystectomy, regardless of the age of the patient. Fewer than one-third of departments perform it routinely.[390] These findings suggest that incidental appendectomy rates may be difficult to estimate with precision at the majority of hospitals where it is not a routine practice.

**Minimum bias.** Since incidental appendectomy appears to be contraindicated in an elderly population, very few (if any) cases would be justified by patients' preoperative characteristics. There are documented cases of discovery of diseased appendices during other abdominal surgery, which would justify incidental appendectomy. However, it is unlikely that the number of diseased appendices found incidental to other surgeries would vary systematically across hospitals. There are no identified risk factors that put an individual at higher risk of having an asymptomatic diseased appendix; therefore, it is impossible to estimate the magnitude of bias.

**Construct validity.** We located no articles explicitly addressing the construct validity of this indicator. Though most of the available evidence appears to contraindicate incidental appendectomy in the elderly, performance of the procedure is subject to patient and surgeon preference. Therefore, incidental appendectomy rates may correlate poorly with other measures of hospital performance. Recent surveys[390] and reviews[387] suggest that incidental appendectomy is still a common practice at some academic centers that enjoy a "high quality" reputation.

**Fosters true quality improvement.** We found no evidence regarding gaming for this indicator. However, since incidental appendectomy does not generally affect hospital payment, widespread use of this indicator may lead to less frequent coding of the procedure, when it is performed. Since removal of an inflamed appendix is clearly appropriate, it seems unlikely that patients would be denied a necessary appendectomy. Of course, a reduction in the rate of incidental appendectomy may lead to a subsequent increase in the incidence of acute appendicitis.

**Prior use.** Incidental appendectomy in the elderly is a provider-level utilization indicator in the current HCUP I indicator set. Otherwise, it has not been widely used as an indicator of quality.

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw provider level rate/standard deviation | 2.7%, 3.5% | |
| Systematic provider-level standard deviation** | 1.9% | Moderate |
| Provider variation as a percentage of total variation** | 1.4% | High |
| Signal ratio | 55.4% | Moderate |
| R-Square** | 67.3% | Moderate |
| **APR-DRG, age-, gender- adjusted | | |
| *Minimum Bias – APR-DRG risk adjustment* | | |
| Signal variance change with risk adjustment | No Change | Good |
| Absolute impact: | | |
| Average absolute change (in %) | 5.0% | Very Good |
| Relative impact: | | |
| Rank correlation | 0.988 | Very Good |
| Percent remaining in high decile/low decile | 82.9% / 94.6% | Good |
| Percent changing more than 2 deciles | 0.3% | Very Good |

**Precision.** This indicator is precise, with a raw provider level mean of 2.7% and a standard deviation of 3.5%. The systematic provider level standard deviation is moderate, at 1.9%. The provider level variation accounts for a high percentage of total variation, at 1.4%. This means that relative to other indicators, a higher percentage of the variation occurs at the provider level, rather than the discharge level, although more of the variation occurs at the discharge level than for some indicators. Finally, the signal ratio is moderate, at 55.4%. This means that it is likely that some of the observed differences do not represent true differences in provider performance. The moderate R-square for this indicator reflects that relatively modest amount of signal that can be extracted using multi-variate techniques, though these techniques do help somewhat relative to univariate techniques.

**Bias.** The indicator performs well to very well on the multiple measures of minimum bias. The rank correlation is high at 0.988. Risk adjustment does not appear to impact the extremes of the distribution substantially. Ninety-four percent of providers in the low decile without risk adjustment remain after risk adjustment; 82.9% remain in the highest decile. The absolute magnitude of the impact of risk adjustment is modest.

**Construct validity.** Incidental appendectomy loads highly on factor 3. It is positively related to cesarean section delivery and negatively related to VBAC.

### *Discussion*

Incidental appendectomy is contraindicated in the elderly population, as this population has both a lower risk for developing appendicitis and a higher risk of postoperative complications when incidental appendectomy is performed. The procedure is not currently performed widely in the elderly, though it is still performed. The contraindications against this procedure in the elderly, as noted in the literature, are compelling and thus this indicator is recommended.

Given the low rate of incidental appendectomies, the precision for this indicator may be lower than other indicators. Our empirical analyses found that this indicator is moderately precisely measured. The moderate signal ratio suggests that some of the observed differences do not reflect true systematic differences in performance.

Empirically we found little evidence of bias in this indicator. The relative effect is somewhat minimal, and thus the bias with respect to provider differences is not likely to be high. As it is unlikely that the only indication for this procedure, the incidental discovery of an asymptomatic diseased appendix, will vary systematically between providers, it is unlikely that this indicator will be substantially biased.

Overall, this indicator is recommended for inclusion in the HCUP II QI set. It received an empirical rating of 13 out of 26, and smoothing is recommended. This indicator is recommended with one caveat of use. As a utilization indicator, the construct validity relies on the actual inappropriate use of procedures in hospitals with high rates, and this should be investigated further.

## INDICATOR 10: BI-LATERAL CARDIAC CATHETERIZATION RATE

| | |
|---|---|
| Indicator | Bi-lateral cardiac catheterization rate |
| Relationship to Quality | Bi-lateral catheterization is contraindicated in most patients without proper indications. As such, lower rates of bi-lateral catheterization represent better quality care. |
| Benchmark | State, regional or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Provider level bi-lateral cardiac catheterizations per 100 discharges with procedure code of heart catheterization. |
| Outcome of Interest | All simultaneous right and left heart catheterizations (see Appendix 6).<br><br>Exclude valid indications for right sided catheterization (see Appendix 6) in any diagnosis field. |
| Population at Risk | All heart catheterizations in any procedure field (see Appendix 6).<br><br>Include only coronary artery disease (see Appendix 6). Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates) |

***Evidence from the literature***

**Face validity.** The diagnostic evaluation of patients with presumptive coronary artery disease often involves cardiac catheterization with coronary angiography. Left-sided catheterization provides very useful information about coronary anatomy, as well as left ventricular function and valvular anatomy. Right-sided catheterization is often performed at the same time, but this practice raises two appropriateness issues. First, without a specific indication for right heart catheterization, the clinical yield is extremely low. In the most rigorous prospective study of this phenomenon, case management was changed for only 1.5% of patients who received an incidental right heart catheterization without a listed indication.[391] Similar results have been reported from two retrospective studies,[392, 393] while other studies have failed to distinguish unsuspected right-sided abnormalities that affected management from those that did not.[394] Second, the marginal cost of right heart catheterization has been estimated to exceed $650 per case and $120 million for the nation.

In response to these research findings, the American College of Cardiology and the American Heart Association published guidelines for cardiac catheterization laboratories stating that "without specific indications, routine right heart catheterizations…are unnecessary."[395] Similar guidelines have been published by other medical and public health organizations, such as the Cardiac Advisory Committee of the New York State Department of Health and the Texas Medical Association's Committee on Cardiovascular Diseases.

**Precision.** In 1996, about 23% of all Medicare beneficiaries who underwent left heart catheterization also underwent right heart catheterization. At the state level, this percentage varied from 11% in Oklahoma to 48% in Massachusetts and 53% in Washington, DC.[396] Given

that more than 1.2 million inpatient cardiac catheterizations were performed in the US in 1998, this measure should be estimable with reasonable precision.[397]

**Minimum bias.** Bilateral cardiac catheterization is considered appropriate in the presence of certain clinical indications: suspected pulmonary hypertension or significant right sided valvular abnormalities, congestive heart failure, cardiomyopathies, congenital heart disease, pericardial disease, and cardiac transplantation. The validity of this measure rests on the assumption that the prevalence of these clinical indications is low and/or relatively uniform across the country. Unfortunately, the true prevalence of these indications cannot be reliably derived from administrative data. However, Malone et al [398] found that substantial variation in the use of bilateral catheterization persisted among 37 cardiologists at two large community hospitals, even after adjusting for clinical indications. Bias is likely to account for an even smaller share of variation at the hospital level.

Another source of potential bias is the large number of catheterizations performed on an outpatient basis. In 1996, 472,000 of 1,633,000 catheterizations were performed on an outpatient basis.[297]

**Construct validity.** We located no articles explicitly addressing the construct validity of this indicator.

**Fosters true quality improvement.** We found no evidence regarding gaming for this indicator. When bilateral cardiac catheterization does not affect hospital payment (as in the DRG system), widespread use of this indicator may lead to less frequent coding of the procedure, when it is performed. It seems unlikely that patients would be denied a bilateral catheterization when the clinical situation clearly warrants it. However, a reduction in the rate of routine bilateral catheterization may lead to rare, but potentially serious, missed diagnoses (e.g., pulmonary hypertension). The long-term significance of missing these rare diagnoses is unclear.

**Prior use.** Bilateral cardiac catheterization has been widely used as an indicator of quality in the Medicare program. It is one of five quality indicators included in the Medicare Quality of Care Report of Surveillance Measures [399]. From 1993 to 1999, Peer Review Organizations in 20 states developed programs to reduce excessive rates of bilateral cardiac catheterization through education and outreach. Ten of these projects have released results; all documented dramatic utilization changes at the targeted hospitals. It has been estimated that these programs averted at least 6,126 unnecessary bilateral catheterizations.[400] Four of these state-based quality improvement projects have been described in the peer-reviewed literature,[401-404] and one documented a spillover effect in the ambulatory setting.[405] The results of these studies suggest that right heart catheterization rates represent an actionable opportunity for quality improvement.

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw provider level rate/standard deviation | 19.3%, 20.0% | |
| Systematic provider-level standard deviation** | 16.1% | Very High |
| Provider variation as a percentage of total variation** | 14.4% | Very High |
| Signal ratio** | 94.3% | Very High |
| R-Square** | 96.2% | Very High |
| **APR-DRG, age-, gender- adjusted | | |
| *Minimum Bias – APR-DRG risk adjustment* | | |

| | | |
|---|---|---|
| Signal variance change with risk adjustment | No change | Good |
| Absolute impact: | | |
| Average absolute change (in %) | 6.1% | Very Good |
| Relative impact: | | |
| Rank correlation | .988 | Very Good |
| Percent remaining in high decile/low decile | 70.7% / 96.6% | Good/ V.G. |
| Percent changing more than 2 deciles | 0.2% | Very Good |

**Precision.** This indicator is very precise, with a raw provider level mean of 19.3% and a substantial standard deviation of 20.0%. The systematic provider level standard deviation is very high, at 16.1%. The provider level variation also accounts for a very high percentage of total variation, at 14.4%. This means that relative to other indicators, a higher percentage of the variation occurs at the provider level, rather than the discharge level. Finally, the signal ratio is very high, at 96.2%. This means that it is likely that the observed differences in provider performance represent true differences in provider performance. The high R-Square demonstrates that a very large proportion of the signal can be extracted using multivariate techniques. However, since the signal ratio is very high to begin with, MSX smoothing adds less additional impact relative to other indicators.

**Bias.** Signal variance does not change with APR-DRG risk adjustment. The indicator performs well on the multiple measures of minimum bias. The rank correlation is very high at 0.988, and risk adjustment does seem to disproportionately impact the extreme high end relative to the extreme low decile, though the impact is still relatively modest. The absolute impact is also minimal.

**Construct validity.** Bilateral catheterization rate loads very highly on factor two. It is positively related to CABG mortality and negatively related to laparoscopic cholecystectomy.

### *Discussion*

Right side coronary catheterization incidental to left side catheterization has little additional benefit for patients without indications of right side catheterization. Despite guidelines that have been set forth discouraging such practice, the practice continues in some hospitals.

This indicator received one of the highest precision ratings. Provider level variation accounts for a relatively large portion of the total variation compared to other indicators, meaning that variation for this indicator is influenced less by discharge level variation (patient level) than total variation for other indicators. Given the very high signal ratio, it is likely that the observed differences in the provider performance represent true differences in provider performance, rather than random variation. Multivariate smoothing techniques do give some additional benefits, though these benefits are modest, so MSX techniques may not be required for this indicator. Univariate smoothing is always recommended, however.

In our analyses of minimum bias, we identified very little bias in this indicator, when adjusting for APR-DRGs. One study of cardiologists found that clinical characteristics did not account for the observed variance between providers. While there are appropriate uses for this indicator, it is unlikely that such indications would vary systematically between providers to the degree necessary to explain the observed variance. Thus, we would expect this indicator to be only minimally biased.

Overall, this indicator is recommended for inclusion in the HCUP II QI set. It received an empirical rating of 25 out of 26. This indicator is recommended with two caveats of use. First, outpatient procedures may result in selection bias for this indicator, and should be examined.

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

Section 3.E. Indicator 10. Bilateral Catheterization Rate.

**INDICATOR 11: SUCCESSFUL VAGINAL BIRTH AFTER CESAREAN SECTION (VBAC) RATE**

| | |
|---|---|
| Indicator | VBAC rate |
| Relationship to Quality | VBAC has been identified as a potentially underused procedure. As such higher appropriate VBAC rates represent better quality care. |
| Benchmark | State, regional or peer group average |

Method:

| | |
|---|---|
| Quality Measure | Provider-level vaginal births per 100 discharges with diagnosis of previous c-section. |
| Outcome of Interest | Number of vaginal births in women with diagnosis of previous C-section (see Appendix 6). |
| Population at Risk | All deliveries with previous C-section diagnosis in any diagnosis field (see Appendix 6). |

***Evidence from the literature***

**Face validity.** The rate of cesarean section (CS) in the United States increased from 5.5% in 1970 to a high of 24.7% in 1988; with a subsequent decrease to decreased to 20.7% in 1996.[339] In the 1980s, the Department of Health and Human Services concluded that encouraging vaginal birth after cesarean section (VBAC) represented a safe way to decrease the overall CS rate [406] and subsequently set the following targets for the Healthy People 2000 Objectives: a CS rate = 15 per 100 births, with a primary CS rate of 12 or fewer per 100 births and an increase in the vaginal birth after CS (VBAC) rate to 35 or more per 100 births. This target, as well a number of studies in the literature suggesting that increasing VBAC rates could be safely achieved,[407-412] led to the common adoption VBAC rates as a QI.

Despite the widespread use of VBAC rates as a QI (including in HCUP I), a randomized trial comparing a trial of labor vs. elective repeat cesarean section has yet to appear. Moreover, while physicians and policy makers have presumed that encouraging increased VBAC rates conforms to patient preferences, approximately one third of patients prefer to pursue elective repeat cesarean section.[413-416] In fact, many physicians appear to consider cesarean delivery preferable to vaginal delivery, given the potential complications of the former.[417] Lastly, a recent article indicates that, accounting for costs and not charges, VBAC is unlikely to achieve significant cost savings compared to repeat CS.[418]

Recommendations for increasing the VBAC rate began to appear in the U.S. in the 1980s. A number of observational studies in the 1980s and early 1990s suggested that a trial of labor in patients with previous cesarean delivery (CD) represented a safe practice.[407-412] More recently, however, evidence for increased risk of maternal and fetal complications associated with this practice has appeared.[365, 419-424] These complications include uterine rupture, maternal infection and hypoxic injury to the fetus.[365] Uterine rupture represents the most serious of these complications, and a recent study by the CDC indicates that administrative data do not adequately capture this complication.[425]

Interestingly, the authors of some studies suggesting an increased risk of uterine rupture associated with a trial of labor have nonetheless concluded that VBAC is overall a safe procedure.[411, 412, 426, 427] Thus, the policy of recommending VBAC represents to some degree a matter of opinion on the relative risks and benefits of a trial of labor in patients with previous CS. Given the substantial more recent evidence regarding maternal and fetal complications as a result of promoting VBAC and the potential for differences of opinion regarding the significance of these adverse outcomes, the existing HCUP I QI may not be regarded as a straightforward corrective for the previous (presumed) overuse of CS. Given the concern regarding maternal and fetal safety, some have advocated for further research to establish the "right rate" of VBAC.[428]

**Precision.** We located no evidence on the precision of this indicator. However, VBAC is a relatively common procedure, and thus we would expect it to be measured with good precision.

**Minimum bias.** In one study,[429] only 42.0% of women with a CS-Vaginal delivery sequence were correctly identified on the second birth certificate as a VBAC. Only 75% of 25,491 women from 1980 through 1988 with a previous cesarean were so designated on the birth certificate; 80% of women with a V- CS sequence were correctly designated as primary cesarean. Although this study employed birth certificates, these findings suggest that administrative data accurately distinguish the mode of current mode of delivery (vaginal vs. CS), but less accurately identify VBAC and primary cesarean delivery.

The proportion of patients with certain sociodemographic profiles[430] and medical indications for CS or contraindications to a trial of labor [358, 360, 431-434] exerts a significant impact on QIs related to rates of CS and VBAC. Although adjusting for case-mix does not eliminate wide variation in rates of CS[435], unadjusted rates only modestly correlate with adjusted rates[358]. Administrative data sources, such as vital records and hospital discharge data, do not include the clinical factors required to identify appropriate candidates for trial of labor. [358, 432, 436, 437] Thus, the denominator for VBAC rates calculated using administrative data will include women with an accepted medical indication for repeat CS delivery (e.g., women with a prior classic CS), not to mention patients who make an informed decision not to pursue a trial of labor.[416]

Prompted by concern over an increase uterine rupture (UR) the Massachusetts Department of Public Health and the CDC conducted an investigation of the validity and reliability of ICD-9-CM codes in hospital discharge data to identify UR cases. The study covered maternal discharges from Massachusetts hospitals from 1990 through 1997; women with and without a history of prior CS were included. Potential cases of UR were identified with an ICD-9-CM diagnostic code in any of the 10 diagnostic fields of 665.0 ("rupture of uterus before onset of labor"), 665.1 ("rupture of uterus during labor," including "rupture of uterus not otherwise specified"), or 674.1 ("disruption of cesarean wound," including "dehiscence or disruption of uterine wound"). Two clinicians then reviewed the medical records of suspected cases to confirm the diagnosis of UR (defined as any unintentional disruption of the uterine wall in a pregnant woman regardless of cause, size, degree of severity, or location).

The design of the study does not permit estimation of the sensitivity of the codes for UR. Positive predictive values (PPVs) were calculated as the number of confirmed cases divided by the number of reviewed suspected cases multiplied by 100. The average PPV during the 8-year period was 50.7% for ICD-9-CM codes 665.0 and 665.1 and 28.6% for code 674.1. The overall PPV of the three codes was 39.8%. Approximately half of the uterine ruptures that result from a trial of labor (among patients with a prior CS) are coded to 674.1 rather than 665.0/1. Furthermore, only about half of the patients coded as having UR (665.0/1) actually had this complication. As an editorial note accompanying the study point out, these coding problems may

reflect the fact that the development of the ICD-9 system predates the interest in monitoring trends in the incidence of UR associated with increased VBAC rates. As ICD-10 codes were also published prior to increased concern over UR, administrative data will have limited use in monitoring the potential negative impacts of attempted VBAC.

**Construct validity.** The likelihood that a patient will undergo VBAC correlates with certain provider and institutional variables,[354, 361, 433, 438, 439] suggesting that certain providers are more likely to adapt to changes in policy or technology.

**Fosters true quality improvement.** Promotion of VBAC a QI has led to successful increases in the VBAC rate in some cases, [440, 441] but not in others.[442] The major opportunity for gaming (i.e., spurious improvement) lies in focusing on patients more likely to successfully complete a trial of labor after previous CS.

**Prior use.** VBAC is one of the most commonly used indicators. In addition to being included in the previous version of the HCUP I indicator set, VBAC is included in and used by JCAHO's IM System, Maryland QI Project,[369] Michigan Hospital Association,[373] HealthGrades.com,[377] Cleveland Health Quality Choice,[374] and the University Hospital Consortium.[370] Further, JCAHO has selected VBAC as one of its core measures.[443]

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw provider level rate/standard deviation | 33.6%, 14.8% | |
| Systematic provider-level standard deviation* | 11.7% | Very High |
| Provider variation as a percentage of total variation* | 5.7% | High |
| Signal ratio* | 83.1% | High |
| R-Square* | 89.5% | High |
| * age- adjusted | | |
| *Minimum Bias – age risk adjustment* | | |
| Signal variance change with risk adjustment | No change | Very Good |
| Absolute impact: | | |
| Average absolute change (in %) | 3.1% | Very Good |
| Relative impact: | | |
| Rank correlation | 0.995 | Very Good |
| Percent remaining in high decile/low decile | 93.1% / 94.3% | V.G./ Good |
| Percent changing more than 2 deciles | 0.0% | Very Good |

**Precision.** This indicator is very precise, with a raw provider level mean of 33.6% and a substantial standard deviation of 14.8%. The systematic provider level standard deviation is very high, at 11.7%. The provider level variation accounts for a high percentage of total variation, at 5.7%. This means that relative to other indicators, a higher percentage of the variation occurs at the provider level, rather than the discharge level, though some variation remains at the discharge level. Finally, the signal ratio is high, at 83.1%. This means that it is likely that the observed differences in provider performance represent true differences in provider performance, although some of the observed difference is due to unobserved differences in patient characteristics. The high R-Square indicates that some additional signal can be extracted using multivariate techniques, though this additional impact is moderate.

**Bias.** Signal variance does not change with risk adjustment. We did not perform APR-DRG risk adjustment because the categories correspond to the outcome of interest. As a result,

the indicator performs very well on the multiple measures of minimum bias, adjusting only for the age of the mother. The rank correlation is very high at 0.995. Age risk adjustment does not seem to impact disproportionately at the extreme high and low end, though some there is some modest impact at the low extreme relative to other indicators. The absolute change is low, and no providers change more than two deciles with risk adjustment.

**Construct validity.** VBAC rate loads very highly on factor three, and is inversely related to cesarean section delivery and to a lesser extent incidental appendectomy.

***Discussion***

Vaginal birth after cesarean section (VBAC) has been implicated as an underused procedure. Healthy People 2010 established a goal of increasing VBAC rates, indirectly by establishing a goal of decreasing cesarean sections in women with previous cesarean section to 63%.

Our empirical tests demonstrated that this indicator is measured with very good precision. Given the high signal ratio, it is likely that the observed differences represent true differences in provider performance, rather than random variation. This cannot fully account for systematic bias in the indicator however. Multivariate smoothing techniques appear to increase modestly the amount of signal that can be extracted. While this amount is high, it remains slightly lower than other indicators.

While we found no change in provider ranking with risk adjustment, this analysis only accounted for the age of the mother, as there are no severity classifications within the APR-DRG for VBAC. The literature review revealed that some clinical factors may contraindicate this procedure, and thus should be risk adjusted for. It is unlikely that these clinical diagnoses, such as previous classic cesarean section, would be available in administrative data. Some clinical information may be available on birth records, and thus linkage to such vital records may provide for better risk adjustment. Further, administrative data does not capture such information as informed preference of the patient or physician in electing repeated cesarean section. Hospitals may vary systematically in the presence of these factors, though this has not been widely studied. Providers may wish to examine case mix to illuminate any potential biases.

We located no other studies that examined the construct validity of VBAC. However, we found that c-section and VBAC are strongly negatively correlated, as one would expect, as increasing the VBAC rate was proposed as one means of reducing c-section rate. However, a simple correlation does not imply causation, nor guarantee that this relationship is evidence of an underlying quality relationship.

Excessive rates of VBAC may increase rates of uterine rupture. Yet, the best rate for VBAC has not been established. We suggest that this measure be used in conjunction with area rates and national rates, and complication rates (maternal uterine rupture and length of stay, neonatal length of stay) to assess whether ones rate is truly too high or too low.

This indicator may indeed encourage better documentation of previous cesarean sections. However, as current evidence indicates that currently previous cesarean section is undercoded, providers with low rates of cesarean section should investigate whether there are substantial undercoding problems.

Overall, this indicator is recommended for inclusion in the HCUP II QI set. It received an empirical rating of 19 out of 26. This indicator is recommended with several caveats of use. First, selection bias due to preferences of patients and other factors may impact performance on this indicator. Second, potential additional bias may result from clinical differences not

identifiable in administrative data, so supplemental risk adjustment with linked birth records or other clinical data may be desirable. Third, as a utilization indicator, the construct validity relies on the actual appropriate use of procedures in hospitals with high rates, and this should be investigated further. Finally, this caution should be maintained for VBAC rates that are drastically below or above the average or recommended rates.

## INDICATOR 12: LAPAROSCOPIC CHOLECYSTECTOMY RATE

| | |
|---|---|
| Indicator | Laparoscopic cholecystectomy rate |
| Relationship to Quality | LC is a new technology with lower risks than open cholecystectomy (removal of the gall bladder). As such a higher rate of LC represents better quality care. |
| Benchmark | State, regional, peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Provider level, Number of laparoscopic cholecystectomies per 100 cholecystectomies |
| Outcome of Interest | Number of laparoscopic cholecystectomies (see Appendix 6). |
| Population at Risk | All non-maternal/non-neonatal discharges age 18 years or older with any procedure code for cholecystectomy (see Appendix 6) in any field.<br><br>Include only discharges with uncomplicated cases: cholecystitis and/or cholelithiasis (see Appendix 6) in any diagnosis field.<br><br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates) |

### *Evidence from the literature*

**Face validity**. Cholecystectomy, surgical removal of the gallbladder, is now generally performed with a laparoscope (in about 75% of uncomplicated cases in the NIS).[444] One of the largest randomized controlled trials (RCTs) comparing cholecystectomy by laparoscopy versus minilaparotomy found that the laparoscopic procedure was associated with less postoperative pain, lower patient-controlled morphine consumption, better postoperative pulmonary function and oxygen saturation, and quicker return to leisure activities, work in the home, and social activities (but no difference in return to employment).[445-451] This study also confirmed that the laparoscopic approach is associated with less postoperative narcotic use, less sick leave, and shorter convalescence (even with conversion rates as high as 13%). Observational studies suggest that laparoscopic surgery also has a lower mortality rate[452-456] and a lower readmission rate.[452-454] Only one large RCT reported no difference in hospital stay, time off work, or return to full activities.[457] The authors attribute this difference to their small-incision (7 cm) minilaparotomy and their unusual blinding method, in which patients were told to start eating, get out of bed, and go home whenever they felt ready, without knowing which procedure they had received.

As a result of these studies, laparoscopic cholecystectomy is now widely accepted as the "standard of care for patients requiring cholecystectomy,"[458] in the absence of specific contra-indications (e.g., coagulopathy, late pregnancy, morbid obesity, cirrhosis). However, this

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

**Precision.** Cholecystectomies are relatively common, so moderately precise estimates of differences in laparoscopic utilization across hospitals can be obtained. In the NIS, the average number of cholecystectomies per hospital is approximately 70, with nearly two-thirds of all hospitals performing at least 3 procedures. Still, random variation in laparoscopic utilization across hospitals in a particular year may be considerable, particularly for hospitals that perform few procedures. Restricting the denominator definition to uncomplicated cases reduces the size of the denominator by about 20%, which further reduces precision.[459]

**Minimum bias**. The current HCUP I definition limits the denominator to uncomplicated cases: non-acute cholecystitis (inflammation of the gallbladder) and/or cholelithiasis (formation of bile stones in the gallbladder). For these patients, cholecystectomy is an elective procedure. Cholecystectomies on patients with acute cholecystitis are performed as emergency procedures, which increases the risk of iatrogenic injury.[460] Higher risks of complications with LC are also associated with older age and the presence of common bile duct stones.[459] Only a limited number of surgeons are comfortable with laparoscopic common bile duct exploration, which is described by the Society of American Gastrointestinal Endoscopic Surgeons as "a complex biliary procedure that demands a well-trained operating room team and facilities and equipment beyond that required for routine LC." In addition, as surgeons become more experienced in LC, they are more likely to perform LC on more difficult patients, such as those with acute cholecystitis[461] and those at older and younger ages.[462, 463] These examples illustrate that patient referral patterns and other selection factors may lead to substantial differences in laparoscopy rates (as a proportion of all cholecystectomies) across hospitals. While many of these patient characteristics can be measured and thus adjusted out in comparisons across hospitals, controlling for other aspects of patient severity may be more difficult.

In 1996, there were an estimated 770,000 cholecystectomies, both open and laparoscopic, in the US – 322,000 ambulatory (42%) and 448,000 inpatients (58%). [297] Thus, use of only inpatient data could be substantially biasing, in that it eliminates those cholecystectomies performed on an outpatient basis, most of which are likely to be laparoscopic.

**Construct validity**. We found no studies explicitly evaluating the construct validity of this indicator. In other words, there is no evidence that hospitals that use the laparoscopic approach more frequently provide better quality of care, according to other measures. One study examined factors associated with the rate of hospital adoption of laparoscopic cholecystectomy in Pennsylvania; the authors reported that "nearly universal adoption occurred at a speed not previously reported" (e.g., within 2 years). Participation in residency training was the only independent predictor of earlier LC adoption.

**Fosters true quality improvement**. This indicator raises several concerns related to perverse incentives. Within 3 years after its introduction, 76% of cholecystectomies in Maryland[462, 463] and 72% of those in Pennsylvania[464] were performed laparoscopically. In both states, and in New York,[465] the advent of laparoscopic surgery led to a substantial (28-34%) increase in the overall cholecystectomy rate, especially involving uncomplicated and elective patients.[466-468] Similar, but less dramatic, increases have been reported from the Veterans Affairs system (10%),[456] Canada (17%), Australia (26%),[469] and Scotland (19%).[470] These trends eliminated most of the expected decline in cholecystectomy-associated deaths in Maryland and in the Veterans Affairs system. The NIH Consensus Development Panel tried to

slow this phenomenon by declaring that "the availability of laparoscopic cholecystectomy should not expand the indications for gallbladder removal."[466] It is not clear whether this declaration has had any moderating effect on community practice. However, providers could readily boost LC rates by recruiting more persons with marginal clinical indications to undergo cholecystectomy.

The other concern in this domain is that the "optimal" LC rate has not been defined. Provider experience may be an important, and desirable, determinant of how the procedure is used across a wide range of clinical situations. At some hospitals, increasing proportionate LC utilization above 75% might lead to more biliary tract or intestinal complications. In addition, previous studies have clearly demonstrated a learning curve, whereby surgeons' outcomes and operating times improve as they gain experience with LC.[471, 472] Technical complication rates appear to stabilize at a low rate after about 75 procedures.[473] Incentives to increase LC utilization may have negative consequences if local physicians lack appropriate training and expertise.

These findings suggest that proportionate LC utilization must be interpreted together with overall cholecystectomy rates (to rule out over-referral of patients who are inappropriate or marginally appropriate) and LC complication or conversion rates (to rule out inappropriate use of LC among patients who would benefit from open surgery).

**Prior use.** The rate of laparoscopic cholecystectomy is a current indicator in the HCUP I QI set. We were unable to find evidence that this measure has been used as a quality indicator in other settings. Indeed, the rapid and nearly universal adoption of LC that occurred between 1990 and 1993 makes future implementation of this measure seem unlikely.

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw provider level rate/standard deviation | 66.2%, 19.2% | |
| Systematic provider-level standard deviation* | 13.3% | Very High |
| Provider variation as a percentage of total variation* | 7.9% | Very High |
| Signal ratio* | 83.1% | High |
| R- Square* | 89.1% | High |
| * age- and gender- adjusted | | |
| *Minimum Bias – age-sex risk adjustment* | | |
| Signal variance change with risk adjustment | No change | Good |
| Absolute impact: | | |
| Average absolute change (in %) | 3.9% | Very Good |
| Relative impact: | | |
| Rank correlation | 0.966 | Very Good |
| Percent remaining in high decile/low decile | 93.8% / 78.8% | V.G./ Fair |
| Percent changing more than 2 deciles | 2.5% | Very Good |

**Precision.** This indicator is very precise, with a raw provider level mean of 66.2% and a substantial standard deviation of 19.2%. The systematic provider level standard deviation is very high, at 13.3%. The provider level variation also accounts for a very high percentage of total variation, at 7.9%. This means that relative to other indicators, a higher percentage of the variation occurs at the provider level, rather than the discharge level. Finally, the signal ratio is high, at 89.1%. This means that it is likely that the observed differences in provider performance

represent true differences in provider performance, although some of the observed differences are due to unobserved differences in patient characteristics. The high R-square demonstrates that a large proportion of signal can be extracted using multivariate techniques, though not as much as some other indicators. The additional impact of multivariate techniques is modest.

**Bias.** Signal variance does not change with risk adjustment. We did not perform APR-DRG risk adjustment because the categories correspond to the outcome of interest. As a result, the indicator performs very well on the multiple measures of minimum bias, using age-sex adjustment. The rank correlation is high at 0.966. Age and sex risk adjustment does seem to disproportionately impact the low extreme relative to the high extreme of the distribution; ninety-four percent of providers in the high decile remain after risk adjustment, while only 78.8 of providers in the low decile remain. The absolute change is low, and relatively few providers change more than two deciles with risk adjustment.

**Construct validity.** Laparoscopic cholecystectomy loads very highly on factor two. It is inversely related to bilateral catheterization, and CABG mortality rate.

***Discussion***

Use of laparoscopic cholecystectomy as opposed to open cholecystectomy is associated with less morbidity in less severe cases. Thus, laparoscopic cholecystectomy has been identified as a potentially underused procedure, when measured as a ratio to total cholecystectomies performed.

The literature review suggests that due to the rapid adoption of LC, there would be very little variation between providers. In contrast, this indicator has a high percentage of variation attributable to providers. This indicates that while many hospitals do use LC, the proportion of LC to all cholecystectomies does in actuality vary between providers.

This indicator performed very well on the empirical test for minimum bias, showing very minimal bias, when adjusting for age and sex. The exception is the impact of providers with low rates; this bias could lead to the misidentification of providers as outliers when in fact they are not. Additional bias was not examined, as there are no severity classifications within the cholecystectomy APR-DRG. However, the literature review does indicate that there may be need to adjust for clinical severity, age, and other factors, since LC may be contraindicated for some patients, and others may not be clinically severe enough to qualify for cholecystectomy at all. The level of clinical detail required may not be available using HCUP data. There is concern that encouraging too many cholecystectomies to be performed laparoscopically could lead to higher complication and conversion to open cholecystectomy rates. Further too many inappropriate procedures in patients without appropriate clinical indications would artificially inflate the laparoscopic cholecystectomy rate without improving quality.

The most troubling source of bias is that up to half of all cholecystectomies are performed on an outpatient basis. This bias decreased the strength of this indicator substantially. Providers should incorporate outpatient data if possible when interpreting this indicator.

While this indicator performed well empirically, the literature review demonstrates some concerns regarding the interpretation and use of this measure. Several steps could be taken when using this measure to ensure appropriate interpretations. Providers should compare rates with area rates. Providers with substantially higher rates than those of providers in the same area may want to assess the appropriateness of LC procedures. Providers with substantially lower rates may want to examine the clinical severity of their patients to assess whether the low rate is due to a more severe case mix, requiring more open cholecystectomies.

Overall, this indicator is recommended for inclusion in the HCUP II QI set. It received an empirical rating of 20 out of 26. This indicator is recommended with several major caveats of use. First, many laparoscopic cholecystectomies are performed on an outpatient basis, and thus may bias this indicator. Second, additional bias may result from clinical differences not identifiable in administrative data, so supplemental risk adjustment using other clinical data may be desirable. Third, as a utilization indicator, the construct validity relies on the actual appropriate use of procedures in hospitals with high rates, and this should be investigated further. Fourth, providers may inflate the laparoscopic cholecystectomy rate by increasing the procedure rate for patients with questionable indications. Finally, caution should be maintained for laparoscopic rates that are drastically below or above the average rates.

## 3.E.3. Area-level utilization measures

**INDICATOR 13: CORONARY ARTERY BYPASS GRAFT (CABG) RATE**

| | |
|---|---|
| Indicator | Area level coronary artery bypass graft (CABG) rate |
| Relationship to Quality | CABG is an elective procedure that may be overused. As such more average rates would represent better quality. |
| Benchmark | State, regional or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Number of CABGs per 100,000 population. |
| Outcome of Interest | Number of CABGs (any procedure field) per 100,000 population (see Appendix 6).<br><br>Age 40 years or older.<br><br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |
| Population at Risk | Population in MSA or county, age 40 years and over. |

***Evidence from the literature***

**Face validity.** Coronary artery bypass graft (CABG) surgery is performed on patients with coronary artery disease (CAD). Most previous studies of small area variation have found relatively high variation in CABG rates, as noted by the systematic component of variation (.758).[474] This systematic component of variation (SCV) "compares geographic variability between DRGs after removing random effects.[474] This variation is not been explained by population characteristics such as age and race.

The clinical indications for CABG in patients with symptoms less major than three-vessel disease, previous myocardial infarction, or less than strongly positive exercise ECG tests are unclear. [262, 475] No randomized controlled trials have demonstrated that CABG improves clinical outcomes in patients with some combination of these indications.

**Precision.** Because adult admissions for CABG are relatively common, it should be possible to generate precise estimates of utilization at the area level. However, random variation in utilization rates may become more problematic for relatively small areas (e.g., zip codes) or underpopulated areas (e.g., rural counties).

**Minimum bias.** Utilization rates standardized at the area level (e.g., adult population of the county or SMSA) may be biased by differences in the prevalence of CAD. The prevalence of CAD may, in turn, be related to the age structure of the population and the prevalence of behavioral or physiologic risk factors such as smoking and hyperlipidemia. Even though race and demographic factors have significant effects on the likelihood of CABG, previous studies have shown that sociodemographic differences account for very little of the observed variation in CABG rates.[262] While one study reports no significant difference between age categories and appropriateness,[262] another reports that "patients 75 years of age and older were more likely to

have surgery classified as uncertain than were younger patients."[475] Although there is some report of variation based on age, it is not likely that this factor explains all the variation observed in CABG rates. Some differences in CABG rates across areas may be attributable to the referral of rural and other patients from outside the area for surgery; however, such referrals are unlikely to explain a large part of the substantial differences in rates across small geographic areas.

**Construct validity.** For this indicator to perform well in identifying true quality of care problems, there must be evidence of significant inappropriate or questionable utilization in population-based studies. In addition, there must be substantial variation in the rate of inappropriate utilization across providers or small areas. Previous studies have classified possible indications for CABG as appropriate, uncertain, or inappropriate, as a method for assessing quality of care.

In a follow-up of a NY appropriateness study, a panel of cardiologists from Duke University reviewed 308 cases. In their results, the rate of inappropriate procedure increased to 6% and the rate of uncertain procedures increased to 12%. [476] In another study of 12 Academic Medical Center Consortium hospitals, the rate of CABG for inappropriate indications was as much as 1.9% overall, ranging from 0 to 5% across hospitals (P=0.02). The rate of CABG for uncertain indications was 7%, ranging from 5 to 8% across hospitals.[475] In 15 randomly selected non-federal hospitals in New York State (which has relatively low CABG rates by US standards), 2.4% of CABG was inappropriate and 7% uncertain. These rates also varied among hospitals.[262]

In 319 randomly selected patients with CABG between July 1987 and June 1988, 16% overall received CABG for inappropriate reasons and 26% received the procedure for uncertain reasons.[317]

In a comparison of Canadian and New York State samples, appropriateness of CABG was measured according to criteria from each nation. Based on both Canadian and US criteria, the Canadian sample held more cases of CABG for uncertain indications. However, based on Canadian criteria, the Canadian sample had fewer cases of CABG performed for inappropriate indications but, based on US criteria, the rate of inappropriate CABG in both samples is lower than that Canadian criteria assess for either of them, and the rates are approximately the same.[477]

In a study of seven of eight public Swedish heart centers, which perform 92% of all bypass surgeries in Sweden, it was found that 9.7% of all CABG procedures were performed for inappropriate indications and 12.3% of procedures were performed for uncertain indications.[478] Finally, among 153 catheterization patients referred to either a university cardiac laboratory or a VA cardiac lab in Maryland, the rate of CABG performed for inappropriate indications ranged from 17 to 46%, based on three different appropriateness criteria: RAND, ACC/AHA, or RAS. The rate of CABG for uncertain indications was 17%, based on RAND criteria, the only set of criteria that accounts for this rating.

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

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

**Prior use.** The hospital-based rate of CABG is a current indicator in the HCUP I QI set. The area-based rate of CABG is a current indicator in the Dartmouth Atlas.[479]

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw area level rate/standard deviation | 180.4, 571.6 | |
| Systematic area-level standard deviation* | 0.25% | Very high |
| Area variation as a percentage of total variation* | 0.21% | Very high |
| Signal ratio* | 97.3% | Very high |
| R-Square* | 97.3% | Very high |
| * age- and gender- adjusted | | |
| *Minimum Bias – age and sex risk adjustment* | | |
| Signal variance change with risk adjustment | No change | Good |
| Absolute impact: | | |
| Average absolute change (in %) | 12.6% | Good |
| Relative impact: | | |
| Rank correlation | 0.654 | Fair |
| Percent remaining in high decile/low decile | 36.4% / 95.5% | Fair/ V.G. |
| Percent changing more than 2 deciles | 35.0% | Fair |

**Precision.** This indicator is moderately precise, with a raw area level mean is 180.4 per 100,000 population and a standard deviation of 571.6. The systematic area level standard deviation is very high, at 0.25%. The area level variation accounts for a very high percentage of total variation, at 0.21%. This means that relative to other indicators, a larger percentage of the variation occurs at the area level, rather than the discharge level. The signal ratio is very high, at 97.3%. This means that it is very likely that the observed differences in area performance represent true differences in area performance. The R-Square is also very high, demonstrating that a large proportion of signal can be extracted with either univariate or multivariate techniques.

**Bias.** Signal variance does not change substantially with risk adjustment. The indicator performed relatively poorly on most measures of minimum bias, using age and sex adjustment. Both the relative and absolute impact of risk adjustment is substantial. The exception is that risk adjustment did not effect the lowest decile substantially, with over 95% of providers in the lowest decile remaining after risk adjustment.

**Construct validity.** CABG rate loads on factor 1, and is related to all other area utilization indicators.

### *Discussion*

CABG was included in the HCUP I QIs as a provider level indicator and a potentially overused procedure. In this report this indicator has been redefined as an area-level indicator. Substantial and systematic small area variation has been noted in the literature. This variation is not explained by sociodemographic characteristics.

This indicator is measured with very high precision. The systematic area-level variation is very high, and accounts for a very high percentage of the total variation. The very high signal ratio suggests that observed variation between areas is likely to reflect true variation in performance. Multivariate techniques do not extract additional signal, but it does not appear to be necessary. Univariate smoothing, as always, is recommended.

In our empirical analysis of minimum bias, risk adjustment by age and sex affected performance of areas substantially. Some clinical factors were also noted in our literature review are appropriate indications for CABG, such as significant coronary artery disease. These factors may be more prevalent in an area with more risk factors, such as smoking, hyperlipidemia, or older age. As such, risk adjustment with demographic data, at minimum, is recommended.

The ideal rate for CABG has not been established and indeed there are cases where CABG is an appropriate and necessary procedure. Several studies have noted that CABG is not often performed for inappropriate indications (under 15%). As such, this indicator is not recommended as a stand alone quality indicator. Rather it is designed for use with volume and mortality indicators. Methods of evaluating the appropriateness of procedures have been established in the literature cited in this report and by other sources. These methods could be used to evaluate the appropriateness of procedures within an area, if this remains a concern.

Area rates are based on the rates for hospitals within an area, and as such do not take into account that some patients are referred into area hospitals from a different area. Examination of data containing patient residence may aid in identifying the extent to which patients are referred into an area. HCUP data may also be used to examine which hospitals contribute the most to the overall area rate.

Overall, this indicator is recommended for inclusion in the HCUP II QI set, inasmuch as it is used in conjunction with other indicators. It received an empirical rating of 19 out of 26. This indicator is recommended with several major caveats of use. As an area utilization indicator, this indicator is a proxy for actual quality problems. This indicator in particular has unclear construct validity, as CABG does not appear to be performed inappropriately often. Finally, caution should be maintained for CABG that are drastically below or above the average or recommended rates.

**INDICATOR 14: HYSTERECTOMY RATE**

| | |
|---|---|
| Indicator | Area level hysterectomy rate |
| Relationship to Quality | Hysterectomy has been identified as a potentially overused procedure. As such, more average rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Number of hysterectomies per 100,000 population . |
| Outcome of Interest | Number of hysterectomies (any procedure field) per 100,000 population (see Appendix 6).<br><br>Females age 18 years and older.<br><br>Exclude discharges with diagnosis for genital cancer, or pelvic or lower abdominal trauma in any diagnosis field. Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |
| Population at Risk | Female population in MSA or county, age 18 years and older. |

***Evidence from the literature***

**Face validity.** Hysterectomy is performed on patients with a number of indications, such as recurrent uterine bleeding, chronic pelvic pain, or menopause, usually in some combination. Small area variation has been noted. One study of variation within the state of Maryland found relatively moderate variation, as noted by the systematic component of variation (.083).[474] This systematic component of variation (SCV) "compares geographic variability between DRGs (diagnosis-related groups) after removing random effects."[474]

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

**Precision.** Because adult admissions for hysterectomy are relatively common, it should be possible to generate precise estimates of utilization at the area level. However, random variation in utilization rates may become more problematic for relatively small areas (e.g., zip codes) or underpopulated areas (e.g., rural counties).

**Minimum bias.** Utilization rates standardized at the hospital level (e.g., all adult discharges) were included in earlier versions of the HCUP I QIs, but are likely to be biased by local market characteristics and referral patterns. In other words, hospitals that specialize in specific services, such as elective or gynecological surgeries, may appear to have higher utilization rates than hospitals that refer such patients elsewhere. Theoretically, utilization rates standardized at the area level (e.g., adult population of the county or SMSA) may be biased by

differences in the prevalence of those indications that warrant the procedure. The prevalence of these indications may, in turn, be related to the age structure of the population and the prevalence of behavioral or physiologic risk factors. Previous studies have shown that observed variation in hysterectomy rates may be accounted for by sociodemographic differences in the patient population. In a study of seven managed care organizations, "older women were more likely than younger women to have received a hysterectomy for appropriate reasons."[263] In a study of women in the UK, "the risk of hysterectomy was significantly related to parity." [480] "Only 5% percent of nulliparous women had a hysterectomy; the risk for women with one to four children varied between 8% and 11%, and the risk for those with five or more children was 31% (P=0.002)."[480] Even after adjusting for parity in a regression model, "the risk of hysterectomy was still significantly associated with educational qualifications."[480]

**Construct validity.** For this indicator to perform well in identifying true quality of care problems, there must be evidence of significant inappropriate utilization in population-based studies. In addition, there must be substantial variation in the rate of inappropriate utilization across providers or small areas. Previous studies have classified possible indications for hysterectomy as appropriate, uncertain, or inappropriate, as a method for assessing quality of care.

In a random sample of 642 hysterectomies (non-emergency and non-oncological), 16% of procedures were deemed inappropriate, based on patient indications, while 25% were deemed uncertain. The rate of inappropriate indications for hysterectomy varied across plans from 10% to 27%.[263]

Another study focused on women receiving hysterectomies in Southern California. The rate of overall inappropriate indications was 70%, varying from 45% to 100% across diagnoses indicative of hysterectomy. Uncertain indications were not evaluated.[481]

**Fosters true quality improvement.** In theory, use of this quality indicator might reduce appropriate as well as inappropriate hysterectomies. Little evidence exists on whether this is likely to occur, or on the extent to which overall hysterectomy rates are correlated with inappropriate hysterectomy rates.

**Prior use.** The hospital-based rate of hysterectomy is a current indicator in the HCUP I QI set. The area-based rate of hysterectomy is a current indicator in the Dartmouth Atlas.[479]

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw area level rate/standard deviation | 419.4, 323.3 | |
| Systematic area-level standard deviation* | 0.19% | Very High |
| Area variation as a percentage of total variation* | 0.10% | High |
| Signal ratio* | 93.6% | Very High |
| R-Square* | 93.7% | Very High |
| * age adjusted | | |
| *Minimum Bias – Age only risk adjustment* | | |
| Signal variance change with risk adjustment | No change | Good |
| Absolute impact: | | |
| Average absolute change (in %) | 3.0% | Very Good |
| Relative impact: | | |
| Rank correlation | 0.996 | Very Good |
| Percent remaining in high decile/low decile | 81.8% / 90.9% | Good |
| Percent changing more than 2 deciles | 0.00% | Very Good |

**Precision.** This indicator is precise, with a raw area level rate of 419.4 per 100,000 population and a substantial standard deviation of 323.3. The systematic area level standard deviation is very high, at 0.19%. The area level variation accounts for a high percentage of total variation, at 0.10%. This means that relative to other area indicators, a higher percentage of the variation occurs at the area level, rather than the discharge level, though some remains at the discharge level. The signal ratio is very high, at 93.6%. This means that it is likely that the observed differences in area performance represent true differences in provider performance. The very high R-square represents the high proportion of signal that can be extracted using multivariate techniques. However, multivariate techniques have little additional impact on the amount of variance that can be extracted from this indicator, since the signal ratio is already very high.

**Bias.** Signal variance does not change substantially with risk adjustment. The indicator performed well on most measures of minimum bias, though the only adjustment was for age. Age adjustment does not impact the absolute ranking substantially. No providers change more than 2 deciles in performance. Risk adjustment appears not to impact the relative rankings of areas substantially.

**Construct validity.** Hysterectomy loads on factor 1, and is related to all other area utilization indicators.

***Discussion***

Hysterectomy has been proposed as a potentially overused procedure. In the HCUP I QIs this indicator was defined with a provider based denominator. It has been redefined in this report as an area-level indicator. Population rates of hysterectomy have been shown to vary systematically by small geographic area, and this variance cannot be explained by systematic clinical or demographic factors. However, patient and physician preference may also play a role in the choice to have a hysterectomy, and thus may affect area rates.

This indicator is measured with good precision. The area-level systematic variation is very high, and accounts for a high percentage of the total variation. The very high signal ratio suggests that observed variation between areas is likely to reflect true variation in performance. Multivariate techniques do not appear to have substantial additional impact on the ability to extract signal from this indicator, though the signal ratio is already very high. Univariate smoothing, as always, is recommended.

In our empirical analysis of minimum bias, risk adjustment with age did not impact area performance substantially. Some clinical factors were noted in our literature review are appropriate indications for hysterectomy. However, it is unlikely that these indications, with the exception of parity, would vary systematically by area, and thus we would not expect this indicator to be substantially biased. Risk adjustment with age is recommended.

The ideal rate for hysterectomy has not been established and indeed there are cases where hysterectomy is an appropriate and necessary procedure. However, several studies have noted relatively high rates of inappropriate indications for surgery (16%-70%). Methods of evaluating the appropriateness of procedures have been established in the literature cited in this report and by other sources. These methods could be used to evaluate the appropriateness of procedures within an area.

Area rates are based on the rates for hospitals within an area, and as such do not take into account that some patients are referred into area hospitals from a different area. Examination of

data containing patient residence may aid in identifying the extent to which patients are referred into an area. HCUP data may also be used to examine which hospitals contribute the most to the overall area rate.

Overall, this indicator is recommended for inclusion in the HCUP II QI set. It received an empirical rating of 22 out of 26. This indicator is recommended with several major caveats of use. As an area utilization indicator, this indicator is a proxy for actual quality problems. This indicator has unclear construct validity, as high utilization of hysterectomy has not been shown to necessarily be associated with higher rates of inappropriate utilization. As only age adjustment is available using administrative data, additional clinical risk adjustment, such as for parity, may be desirable. Finally, caution should be maintained for hysterectomy rates that are drastically below or above the average or recommended rates.

**INDICATOR 15: LAMINECTOMY AND/OR SPINAL FUSION RATE**

| | |
|---|---|
| Indicator | Area level laminectomy and/or spinal fusion rate. |
| Relationship to Quality | Laminectomy has been identified as a potentially overused procedure. As such, more average rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Number of laminectomies and/or spinal fusions per 100,000 population. |
| Outcome of Interest | Number of laminectomies and/or spinal fusions (any procedure field) per 100,000 population (see Appendix 6).<br><br>Age 18 years and older.<br><br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |
| Population at Risk | Population in MSA or county, age 18 years and over. |

***Evidence from the literature***

**Face validity.** Laminectomy is performed on patients with a herniated disc or spinal stenosis. Most previous studies of small area variation have found relatively high variation in laminectomy rates, as noted by the systematic component of variation (.292).[474] This systematic component of variation (SCV) "compares geographic variability between DRGs [diagnosis-related groups] after removing random effects.[474] Another study in Maryland found a 0.50 variation rate [(sample standard deviation)/(sample mean)] in laminectomy for selected Maryland service areas.[482] A study of Boston-area hospitals found a 2.2-fold variation in laminectomy rates among districts.[483] Larequi-Lauber et al. report that, in the United States, "the use of back surgery varies from one area to another by as mush as 15-fold."[484] This high variation has not been explained by population characteristics such as age and socioeconomic status.

The clinical indications for laminectomy in patients with minor neurological findings, lengthy restricted activity, and equivocal imaging for discal hernia or spinal stenosis are unclear. No randomized controlled trials have demonstrated that laminectomy improves clinical outcomes in patients with these uncertain indications.

**Precision.** Because adult admissions for laminectomy are relatively common, it should be possible to generate precise estimates of utilization at the area level. Indeed, even after accounting for random variations, large differences in area laminectomy rates remain.[474] However, random variation in utilization rates may become more problematic for relatively small areas (e.g., zip codes) or sparsely populated areas (e.g., rural counties).

**Minimum bias.** Utilization rates standardized at the area level (e.g., adult population of the county or SMSA) may be biased by differences in the prevalence of herniated disc or spinal

stenosis. The prevalence of herniated disc or spinal stenosis may, in turn, be related to the age structure of the population and the prevalence of behavioral or physiologic risk factors. However, previous studies have shown that sociodemographic differences and other measurable population characteristics account for very little or none of the observed variation in laminectomy rates.[482]

**Construct validity.** For this indicator to perform well in identifying true quality of care problems, there must be evidence of significant inappropriate utilization in population-based studies. In addition, there must be substantial variation in the rate of inappropriate utilization across providers or small areas. Previous studies have classified possible indications for laminectomy as appropriate, uncertain, or inappropriate, as a method for assessing quality of care.

In an assessment of cases at one Swiss hospital, 23% of patients received surgical treatment for herniated discs for inappropriate reasons and 29% received surgical treatment for uncertain indications.[485] In another study of teaching hospital patients undergoing surgery for herniated disc or spinal stenosis (lumbar discectomy or spinal stenosis surgery), 38% of surgeries were performed for inappropriate indications. Uncertain indications were combined with "appropriate" category.

**Fosters true quality improvement.** Little evidence exists on whether use of area laminectomy rates as a quality indicator would lead to less performance of laminectomies for inappropriate or uncertain indications, without reducing the use of laminectomy for appropriate indications.

**Prior use.** The hospital-based rate of laminectomy is a current indicator in the HCUP I QI set. The area-based rate of laminectomy is a current indicator in the Dartmouth Atlas.[479]

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw area level rate/standard deviation | 139.0, 347.5 | |
| Systematic area-level standard deviation* | 0.13% | High |
| Area variation as a percentage of total variation* | 0.10% | Very High |
| Signal ratio* | 96.7% | Very High |
| R-Square* | 96.7% | Very High |
| * age- and gender- adjusted | | |
| *Minimum Bias – age-sex risk adjustment* | | |
| Signal variance change with risk adjustment | No change | Good |
| Absolute impact: | | |
| Average absolute change (in %) | 6.3% | Very Good |
| Relative impact: | | |
| Rank correlation | 0.933 | Good |
| Percent remaining in high decile/low decile | 31.8% / 95.5% | Fair/ Good |
| Percent changing more than 2 deciles | 7.4% | Good |

**Precision.** This indicator is moderately precise, with a raw area level mean of 139.0 per 100,000 population and a standard deviation of 347.5. The systematic area level standard deviation is high, at 0.13%. The area level variation accounts for a very high percentage of total variation, at 0.10%. The signal ratio is very high, at 96.7%. This means that it is very likely that the observed differences in area performance represent true differences in area performance.

Multivariate smoothing does not have additional impact in extracting signal, mainly due to the already very high ratio.

**Bias.** The signal variance does not change with age-sex risk adjustment. The indicator performed fairly to well on most measures of minimum bias. Risk adjustment does not impact the absolute performance of areas substantially (6.3%). Only 7.4% of areas changed more than 2 deciles. Risk adjustment appears to impact the highest decile disproportionately to the lowest decile, with 31.8 percent of areas in the highest decile remaining after risk adjustment, compared to 95.5% in the lowest decile. However, the overall relative impact of risk adjustment was moderate, with a high rank correlation of .933.

**Construct validity.** Laminectomy rate loads on factor 1, and is related to all other area utilization indicators.

***Discussion***

Laminectomy has been proposed as a potentially overused procedure. This indicator was defined with a provider level denominator in the HCUP I QIs. This report recommends its use as an area level indicator. Laminectomy has been shown to vary widely and systematically between areas. Sociodemographic or clinical factors do not explain this variation. However, patient and physician preference may also play a role in the choice to have a laminectomy, which may in turn affect area rates.

This indicator is measured with high precision. The area-level systematic variation is high, and accounts for a very high percentage of the total variation. The very high signal ratio suggests that observed variation between areas is likely to reflect true variation in performance.

In our empirical analysis of minimum bias, risk adjustment by age and sex only minimally affected the performance of providers, suggesting that performance is not highly influence by the demographic breakdown of the population. In addition, risk adjustment appears to affect areas with the highest rates substantially. This means that without adequate risk adjustment areas may be mislabeled as outliers when in fact they are not. Some clinical factors were noted in our literature review are appropriate indications for laminectomy. However, it is unlikely that these indications would vary systematically by area, and thus we would not expect this indicator to be substantially biased. Risk adjustment for age and sex is recommended.

The ideal rate for laminectomy has not been established and indeed there are cases where laminectomy is an appropriate and necessary procedure. However, several studies have noted relatively high rates of inappropriate procedures (23%-38%). Methods of evaluating the appropriateness of procedures have been established in the literature cited in this report and by other sources. These methods could be used to evaluate the appropriateness of procedures within an area.

High area rates are based on the rates for hospitals within an area, and as such do not take into account that some patients are referred into an area hospitals from a different area. Examination of data with patient residence can aid in illuminating the extent to which patients are referred into an area. HCUP data may be used to examine which hospitals contribute the most to the overall area rate.

Overall, this indicator is recommended for inclusion in the HCUP II QI set. It received an empirical rating of 20 out of 26. This indicator is recommended with several major caveats of use. As an area utilization indicator, this indicator is a proxy for actual quality problems. This indicator has unclear construct validity, as high utilization of laminectomy has not been shown to necessarily be associated with higher rates of inappropriate utilization. Finally, caution should be

maintained for laminectomy rates that are drastically below or above the average or recommended rates.

## INDICATOR 16: PERCUTANEOUS TRANSLUMINAL CORONARY ANGIOPLASTY (PTCA) RATE

| | |
|---|---|
| Indicator | Area level PTCA rate. |
| Relationship to Quality | PTCA has been identified as a potentially overused procedure. As such, more average rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Number of PTCA procedures per 100,000 population. |
| Outcome of Interest | Number of PTCA procedures (any procedure field) per 100,000 population (see Appendix 6).<br><br>Age 40 years and over.<br><br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |
| Population at Risk | Population in MSA or county, age 40 years and over. |

***Evidence from the literature***

**Face validity.** Percutaneous transluminal coronary angioplasty (PTCA) is performed on patients with coronary artery disease (CAD). Previous studies of small area variation have found substantial variation in PTCA rates, though most analyses have been performed on Medicare data.[486]

The clinical benefit of PTCA in many patients with unstable or chronic angina and recent myocardial infarction (MI) is unclear. No randomized controlled trials have demonstrated that PTCA improves clinical outcomes in many patients who commonly receive the procedure, and previous studies have documented large differences across hospitals in the likelihood of treatment with PTCA after MI and in other clinical settings.

**Precision.** Because adult admissions for PTCA are relatively common, it should be possible to generate precise estimates of utilization at the area level. However, random variation in utilization rates may become more problematic for relatively small areas (e.g., zip codes) or underpopulated areas (e.g., rural counties).

**Minimum bias.** Utilization rates standardized at the area level (e.g., adult population of a county or SMSA) may be biased by differences in the prevalence of coronary artery disease. The prevalence of CAD may, in turn, be related to the age structure of the population and the prevalence of behavioral or physiologic risk factors such as smoking and hyperlipidemia. Many previous studies have also shown that sociodemographic differences affect PTCA utilization, e.g., black patients are significantly less likely to undergo PTCA than non-black patients, particularly when indications were deemed equivocal. Little evidence exists on the extent to which area differences in socioeconomic and clinical characteristics may explain area differences in PTCA rates, though the large variations in rates across small geographic areas suggest that population characteristics are unlikely to explain most of the differences.[487]

A second potential source of bias is the performance of procedures on an outpatient basis, which would not be captured in HCUP data and thus would not be reflected in HCUP-based area rate measures. However, less than 10% of PTCAs were performed on an outpatient basis in recent years, and so outpatient procedures do not appear to be quantitatively important enough to explain much of the observed variation.[297]

**Construct validity.** For this indicator to perform well in identifying true quality of care problems, there must be evidence of significant inappropriate utilization in population-based studies. In addition, there must be substantial variation in the rate of inappropriate utilization across providers or small areas. Previous studies have classified possible indications for PTCA as appropriate, uncertain, or inappropriate, as a method for assessing quality of care.

In a study of seven of eight Swedish public heart centers, 38.3% of all PTCA procedures were performed for inappropriate indications and 30% for uncertain indications.[478] Another study of cardiac catheterization labs in Maryland used three different sets of criteria (RAND, ACC/AHA, and RAS) to assess the appropriateness of PTCA in these labs. Inappropriate indications ranged from 22 to 49%. The rate of uncertain indications was 29%, according to RAND criteria, the only set that includes this rating. [487] In a follow-up study of a coronary angiography study conducted in New York, a panel of Duke University cardiologists reviewed 308 records for appropriateness. Rates of PTCA performed for both inappropriate and uncertain reasons increased. That for in inappropriate indications was raised to 12%, while the rate of procedures performed for uncertain indications became 27%. [476]

**Fosters true quality improvement.** In an effort to raise public perception of its quality without actually doing so, service providers might attempt to game results, or take action that would significantly decrease the incidence of an indicator or increase provider risk for these cases. This leads to an inaccurate assessment of quality by the public. Ideally, gaming is avoided in practice, present only in theory. However, providers might engage in practices, such as miscoding cases or recruiting patient groups that are known to have increased risk of CAD, in order to achieve more favorable quality assessment results with regard to PTCA. Perhaps instead of serving as quality assessments, results of these appropriateness studies can serve as guidelines for difficult clinical decisions. Patients and their physician might use them to spark questions and discussion about CAD, the patient's specific indications, and the most appropriate treatment, one that poses the least risk to the patient. [488] Implementing the results of these studies on an individual level might relieve undue stress of providers and reduce the temptation to inaccurately depict provider quality.

**Prior use.** The area-based rate of PTCA is a current indicator in the Dartmouth Atlas.[479]

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw area level rate/standard deviation | 190.8, 455.6 | |
| Systematic area-level standard deviation* | 0.28% | Very High |
| Area variation as a percentage of total variation* | 0.21% | Very High |
| Signal ratio* | 97.3% | Very High |
| R-Square* | 97.3% | Very High |
| * age- and gender- adjusted | | |
| *Minimum Bias – age-sex risk adjustment* | | |
| Signal variance change with risk adjustment | No change | Good |
| Absolute impact: | | |

| | | |
|---|---|---|
| Average absolute change (in %) | 10.4% | Good |
| Relative impact: | | |
| Rank correlation | 0.671 | Fair |
| Percent remaining in high decile/low decile | 36.4% / 95.5% | Fair/ V.G. |
| Percent changing more than 2 deciles | 35.5% | Fair |

**Precision.** This indicator is precise, with a raw area level mean of 190.8 per 100,000 population and a standard deviation of 455.6. The systematic area level standard deviation is very high, at 0.28%. The area level variation also accounts for a very high percentage of total variation, at 0.21%. This means that relative to other indicators, a higher percentage of the variation occurs at the area level, rather than the discharge level. The signal ratio is also very high, at 97.3%. This means that it is very likely that the observed differences in area performance represent true differences in area performance. The very high R-square represents the high proportion of signal that can be extracted using multivariate techniques. However, multivariate methods have no additional impact, due to the already very high signal ratio.

**Bias.** Signal variance does not change substantially with risk adjustment. The indicator performs poorly on most measures of minimum bias. Risk adjustment impacts the absolute performance of areas moderately. In addition, 35.5% of areas changed more than 2 deciles. Risk adjustment appears to impact the highest decile disproportionately to the lowest decile, with 36.4 percent of areas in the highest decile remaining after risk adjustment, compared to 95.5% in the lowest decile. The overall relative impact of risk adjustment was also substantial, with a low rank correlation of .671.

**Construct validity.** PTCA rate loads on factor 1, and is related to all other area utilization indicators.

***Discussion***

The appropriateness and potential overuse of PTCA has been discussed in the literature. PTCA rates have been shown to vary widely and systematically between areas. Sociodemographic differences do not explain all of this variation, though inappropriate indications do not account fully for the variance either. Patient and physician preference may also play a role in the variation of PTCA rates by area.

Given the frequency that this procedure is performed, it is expected to be precise. However, some hospitals may perform very few procedures; precision may be a particular problem for these hospitals. Our empirical analysis found the precision to be very high, with very high systematic variation. The very high signal ratio suggests that any observed differences are likely to reflect true differences in performance. Multivariate smoothing techniques do not appear to have additional impact on the ability to extract signal for this indicator, primarily due to the already high signal ratio. Univariate smoothing, as always, is recommended.

In our empirical analysis of minimum bias, risk adjustment by age and sex does affect the performance of providers. In addition, risk adjustment appears to affect areas with the highest rates substantially. This means that without adequate risk adjustment areas may be mislabeled as outliers when in fact they are not. Some clinical factors were noted in our literature review are appropriate indications for PTCA. These factors may be related to the prevalence of coronary artery disease and as such may be more prevalent in areas with an older age structure or higher rates of smoking or hyperlipidemia. It is unlikely that the area variance resulting from these factors would be substantial enough to account for all the observed variance, and thus we would not expect this indicator to substantially biased. Risk adjustment by age and sex is recommended.

A significant, but small about of PTCA procedures are performed on an outpatient basis (<10%), potentially biasing this indicator. If available, areas may wish to examine both inpatient and outpatient rates together.

The ideal rate for PTCA has not been established and indeed there are cases where PTCA is an appropriate and necessary procedure. However, several studies have noted relatively high rates of inappropriate procedures (up to 48%). Other studies report very low rates of inappropriate use, leaving the question of the validity of this indicator open for interpretation. Methods of evaluating the appropriateness of procedures have been established in the literature cited in this report and by other sources. These methods could be used to evaluate the appropriateness of procedures within an area.

Area rates are based on the rates for hospitals within an area, and as such do not take into account that some patients are referred into area hospitals from a different area. Examination of data containing patient residence may aid in identifying the extent to which patients are referred into an area. HCUP data may also be used to examine which hospitals contribute the most to the overall area rate.

Overall, this indicator is recommended for inclusion in the HCUP II QI set, though it is recommended only for use with measures of mortality and volume. It received an empirical rating of 19 out of 26. This indicator is recommended with several major caveats of use. As an area utilization indicator, this indicator is a proxy for actual quality problems. This indicator has unclear construct validity, as high utilization of PTCA has not been shown to necessarily be associated with higher rates of inappropriate utilization. A minor source of bias may be the small amount of procedures performed on an outpatient basis. Finally, caution should be maintained for PTCA rates that are drastically below or above the average or recommended rates.

## 3.E.4. Ambulatory Care Sensitive Condition Measures

The literature review of the evidence related to potentially avoidable hospital admissions is limited for each indicator because many of the indicators have been developed as parts of sets. Therefore, prior to relating evidence on specific indicators, we introduce the strategy for applying the evaluation framework to potentially avoidable hospital admissions, and provide general information applicable to all indicators in this category.

Only five studies[276, 284, 489-491] have attempted to validate individual indicators rather than whole measure sets. Hence, a major limitation of this literature is that we know relatively little about which components represent the strongest measures of access and quality. Most of the 5 papers that did report about individual indicators also used a single variable, such as median area-specific income or rural residence, for construct validation. All but one of these papers (Bindman) included adjustment only for demographic factors (e.g., age, sex, race).

Bindman et al.[284] provides the best evidence of validity for individual indicators, by demonstrating strong independent associations between self-rated access to care and hospitalization rates for asthma, CHF, and diabetes (after adjusting for sociodemographic factors, severity of illness, and propensity to admit). In addition, associations for COPD and hypertension were weaker but also statistically significant.

The other four studies have validated a similar wider set of indicators. Billings et al.[489] validated their version of all 5 of these indicators, except hypertension, as well as epilepsy and convulsions, severe ENT infections, bacterial pneumonia, angina, cellulitis, gastroenteritis, kidney/UTI, dehydration, iron deficiency anemia, and PID (with the weakest associations involving COPD and gastroenteritis). Millman and the Institute of Medicine[491] validated the same set of indicators as Billings et al., with the exception of iron deficiency anemia and PID, but with the addition of hypertension and hypoglycemia. Of the validated indicators, the weakest association again involved gastroenteritis. Weissman et al.[276] found the strongest evidence of validity for cellulitis, diabetes, gangrene, malignant hypertension, and pneumonia; with mixed evidence for asthma, CHF, hypokalemia, immunizable conditions, pyelonephritis, and bleeding ulcer. Only one of Weissman's indicators, ruptured appendix, was not validated, though a modified version of this indicator was validated in two of the other studies mentioned above.

Twenty-nine other studies have examined sets of ACSC conditions, without examining indicators independently. The indicators examined by these studies, and details regarding study design are summarized in Appendix 8. Since a majority of the evidence for ACSC indicators comes from studies of sets of ACSC indicators together, this report considers this evidence in addition to the limited evidence for each indicator. The evidence regarding ACSC indicators in general is reported below.

Face validity.

The following questions are addressed separately for each indicator:

1. Have clinical trials demonstrated that specific outpatient therapies can reduce the risk of hospitalization?
2. Have observational studies shown associations between specific outpatient therapies and the risk of hospitalization?
3. Is there general consensus that hospitalizations for this condition are often avoidable or preventable, if the patient has timely access to high-quality outpatient care?

**Precision.** The precision of avoidable hospitalization rates is likely to depend on the size of the denominator. For example, Bindman et al. compared the correlation of preventable hospitalization rates for 5 adult conditions (asthma, COPD, CHF, diabetes, hypertension) between consecutive years, at the level of contiguous zip code clusters, or medical service study areas. The correlation across 250 urban clusters, with a median population of 52,000, was 0.96. By contrast, the correlation across 144 rural clusters, with a median population of 16,000, was only 0.81.[492] Across about 160 zip codes in New York City, aggregate ambulatory-care sensitive (ACS) admission rates in 1993 and 1982 were highly correlated (r=0.92).[283] Finally, a recent abstract from the United Kingdom reported (without supporting data) that "Ambulatory Sensitive Condition" admission rates for Health Authorities, which have a mean population of about 250,000, are "quite stable between subsequent years, as expressed in terms of the Spearman's rank correlation".[493]

**Minimum bias.** Previous studies have documented several characteristics that are associated with either the risk of an avoidable hospitalization (at the individual level) or the avoidable hospitalization rate (at the area level). These factors are potential confounders, or sources of bias, when avoidable hospitalization rates are used as a measure of access to care. Bindman et al. found that condition prevalence was an especially important correlate of hospitalization rates for 5 chronic conditions at the zip code cluster level.[284] Race and measures of socioeconomic status (percent with no college, percent with income <$15,000) were also independently associated with "preventable hospitalization rates," but these factors might be measuring subtle aspects of access to care. Interestingly, propensity to seek care and physician practice style explained very little of the variability in area hospitalization rates. At the individual level, self-reported health status, functional limitations, several chronic diseases (e.g., coronary artery disease, diabetes), and a chronic disease risk score are associated with preventable hospitalizations among Medicare beneficiaries.[280] Adding these factors to a predictive model substantially reduces the effect of income (e.g., adjusted Odds Ratio for income <$9,517 decreases from 1.96 to 1.38).[282]

Propensity to seek care and physician practice style have been evaluated indirectly; Weissman et al. found that mean comorbidity (DRGSCALE) scores based on hospital discharge abstracts differed by insurance status. Lower mean comorbidity scores among uninsured patients for about half of the avoidable hospital conditions might indicate that physicians treating this population apply a lower admission threshold.[276] Clearer evidence of this possible practice pattern comes from a study that demonstrated striking differences in the severity-of-illness of patients hospitalized with diabetes across counties with low, medium, or high hospitalization rates.[494] Hospitals in low-rate counties had a higher proportion of severe admissions than hospitals in high-rate counties (38-46% versus 20%). A survey of patients hospitalized in Nebraska market areas with high (versus low) ACS admission rates suggests that patient delays

in seeking outpatient care, rather than poor access, may be the primary cause of this difference.[495]

**Construct validity.** Most previous studies have assessed the validity of an entire set of avoidable hospital conditions, rather than each condition alone, and have used socioeconomic status as a marker of access to care. These studies have repeatedly shown strong correlations between household income and avoidable hospitalizations, both at the individual level and the area level. At the zip code level, income alone explains 51-84% of the variability in ACS admission rates across 15 metropolitan areas in the US.[283] This association is substantially weaker among persons 65 or more years of age,[275, 281] as one would expect if it is driven by access to care rather than underlying social factors. Avoidable hospitalization rates are higher among uninsured or Medicaid-enrolled persons than among privately insured persons, even after adjustment for race and income.[276] Maternal education was the dominant socioeconomic correlate of "discretionary" hospitalization rates among infants (<2 years), accounting for 89% of total variability across 37 zip code areas.[287]

Fewer studies have tested true measures of access to care. In the best of these studies, Bindman and colleagues[284] showed that a 5-point scale of self-reported "difficulty in receiving medical care when needed" explained 50% of the variability in hospitalization rates for 5 chronic medical conditions (asthma, CHF, COPD, diabetes, and hypertension) across 41 urban zip code clusters in California. Adjustment for condition prevalence, propensity to seek care, physician admitting style, and ecological measures of income, education, insurance, race, and gender, had little effect on the association. By condition, the univariate R-squared was 0.47 for asthma, 0.50 for CHF, 0.27 for COPD, 0.46 for diabetes, and 0.22 for hypertension. Having a regular source of care, and primary care physician/population ratios, were also independently associated with avoidable hospitalization rates, when substituted for self-reported access.[496] These relationships did **not** hold in two separate studies of rural zip codes, suggesting that avoidable hospitalization rates are invalid indicators of access in rural areas.[279, 492]

In other studies, the physician/population ratio for family and general physicians has been more strongly associated with avoidable hospitalization rates than measures that include internists, pediatricians, or all physicians.[286, 497] In studies of Medicaid populations, provider continuity in ambulatory care[498] and usual care received from a community health center[499] were associated with lower avoidable hospitalization rates. However, having a regular source of care (for more than 50% of physician office visits) was not associated with lower avoidable hospitalization rates.[500] Not having a primary care physician was significantly associated with ACS hospitalization (versus non-ACS hospitalization) in South Carolina, after adjustment for socioeconomic factors.[501]

Several studies of Medicare beneficiaries have shown weak and inconsistent associations between access indicators and avoidable hospitalization rates. For example, persons in the Medicare Current Beneficiary Survey who reported problems obtaining health care, or lived in a health professional shortage area, were not at increased risk of preventable hospitalization.[280] Instead, their risk was heavily influenced by "need characteristics" (clinical factors). However, beneficiaries in fair or poor health reportedly were at increased risk if they lived in a **primary care** shortage area.[285] An area-level analysis based on Medicare claims suggests that the association between ACS admission rates and physician/population ratios is limited to the 10% of health care service areas with the most severe shortage of physicians (e.g., <0.628 physicians/1,000 population).[502]

**Fosters real quality improvement.** There is limited evidence regarding whether avoidable hospitalization rates can be decreased through interventions to improve access to or quality of care. Three studies have reported on the impact of recent changes in Medicaid eligibility criteria and program design. Kaestner et al. found no narrowing of the differences in "discretionary" infant (<2 year) hospitalization rates between low, middle, and high-income zip codes, during a period of substantial Medicaid eligibility expansion (1988-1992).[503] Implementation of a Medicaid managed care gatekeeper system in Maryland,(Ref. 81) with fee-for-service reimbursement of designated primary medical providers and 24-hour access, modestly reduced the risk of any hospitalization (OR=0.89) but did not decrease the risk of an "ambulatory care sensitive" hospitalization (OR=0.96). Similar programs in Florida and New Mexico may have led to "significant, but small, reductions" in ACS hospitalization rates for children, although no specific data have been reported.[277] In a cross-sectional study on the impact of physician economic incentives, a capitated group practice achieved far lower ACS hospitalization rates than physicians who participated in 3 independent practice associations or treated patients with indemnity insurance(0.8/1,000 versus 2.7/1,000 and 2.9/1,000, respectively).[288] Establishing new community-based outpatient clinics did not decrease preventable hospitalization rates across the Department of Veterans Affairs' primary service areas.[504]

Because the optimal hospitalization rate for ambulatory care sensitive conditions has not been defined, providers may decrease their rates by failing to hospitalize patients who would truly benefit from inpatient care, by eliminating or discouraging product lines important to the community, or by hospitalizing marginally appropriate patients with other conditions (to inflate the denominator). Although these concerns cannot be dismissed, there is no published evidence of worse health outcomes in association with reduced hospitalization rates for these conditions. However, such evidence has been presented for external hernia, appendicitis, and uterine fibroids.[505]

## INDICATOR 17: ACSC: DEHYDRATION ADMISSION RATE

| | |
|---|---|
| Indicator | Area level admission rate for dehydration. |
| Relationship to Quality | Proper outpatient treatment may reduce admissions for dehydration. As such lower rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Admissions for dehydration per 100,000 population. |
| Outcome of Interest | Discharges with ICD-9 principal diagnosis code for hypovolemia (276.5) per 100,000 population (see Appendix 6).<br><br>Age less than 65 years.*<br><br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |
| Population at Risk | Population in MSA or county, age less than 65.* |

* Rate can also be calculated for age 65 and older.

***Evidence from the literature***

**Face validity:** Dehydration is a serious acute condition that occurs in frail patients and patients with other underlying illnesses, following insufficient attention and support for fluid intake. It is treatable with oral rehydration therapy and/or IV fluids. If left untreated in older adults, serious complications including mortality (over 50%) is very high.[506] Thus, prevention and effective management of dehydration may lead to fewer complications associated with severe dehydration, including hospitalization.

**Precision:** Dehydration is a somewhat common cause of hospital admission. We found little evidence on the precision of this indicator. One study did note that dehydration accounted for 7.3% of total admissions for ACSCs.[282]

**Minimum bias:** We found no literature on the potential bias of this indicator. It is possible that the age structure of the population may affect admission rates for this condition, as the elderly and very young are more susceptible to dehydration. Socioeconomic factors may also affect admission rates, though we found no specific evidence confirming this hypothesis. We found no evidence on how comorbidities or other risk factors that may vary systematically by area influence hospitalization rates for dehydration. Finally, different thresholds for admission of patients with dehydration, rather than differences in quality of care, may also lead to area rate differences.

**Construct validity:** We found little literature on admission for dehydration as an ambulatory care sensitive condition indicator. Dehydration was originally included in John Billings'[489] set of indicators developed for the United Hospital Fund of New York. This set was developed by a physician panel. Evidence on sets of ambulatory care sensitive condition

indicators are summarized at the beginning of this section, and should be referred to for this indicator.

Two studies of ACSC indicators reported validation work for dehydration independent of measure sets. Millman et al.[491] reported that low-income zip codes had 2.1 times more dehydration hospitalizations per capita than high-income zip codes in 11 states in 1988. Billings et al.[489] found that low-income zip codes in New York City (where at least 60% of households earned less than $15,000 in 1988, based on adjusted 1980 Census data) had 2.0 times more dehydration hospitalizations per capita than high-income zip codes (where less than 17.5% of households earned less than $15,000). Household income explained 42% of the variation in dehydration hospitalization rates at the zip code level.

**Fosters true quality improvement:** We found no evidence of the impact such a measure would have on quality. As some dehydration can be managed on an outpatient basis, it is possible that a shift to outpatient care may occur. It is unknown whether hospitalizations for dehydration are appropriate, and thus whether the shift to outpatient care would be appropriate.

**Prior use:** This indicator was included in Billings set of Ambulatory Care Sensitive Conditions, developed in conjunction with the United Hospital Fund of New York.[489]

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw area level rate/standard deviation | 139.9, 103.2 | |
| Systematic area-level standard deviation* | 0.04% | Moderate |
| Area variation as a percentage of total variation* | 0.02% | Moderate |
| Signal ratio* | 88.5% | High |
| R-Square* | 88.9% | High |
| * age- and gender- adjusted | | |
| *Minimum Bias – age-sex risk adjustment* | | |
| Signal variance change with risk adjustment | No change | Good |
| Absolute impact: | | |
| Average absolute change (in %) | 9.2% | Very Good |
| Relative impact: | | |
| Rank correlation | 0.957 | Very Good |
| Percent remaining in high decile/low decile | 81.8% / 90.9% | Good |
| Percent changing more than 2 deciles | 4.6% | Very Good |

**Precision.** This indicator is precise, with a raw area level rate of 139.9 per 100,000 population and a standard deviation of 103.2. The systematic area level standard deviation is moderate, at 0.04%. The area level variation also accounts for a moderate percentage of total variation, at 0.02%. The signal ratio is high, at 88.5%. This means that it is likely that the observed differences in area performance represent true differences in area performance. The high R-square represents the relatively high proportion of signal that can be extracted using multivariate techniques, though this is less than for other indicators. Multivariate techniques appear to have little additional impact.

**Bias.** Signal variance does not change substantially with risk adjustment. The indicator performs well on multiple measures of minimum bias, using age-sex adjustment. The rank correlation is very high at .957. Risk adjustment does not affect lowest decile disproportionately to the highest decile. In addition, few providers change more than 2 deciles with risk adjustment

(4.6%). The absolute magnitude of the impact is minimal, with provider performance changing an average 9.2% relative to the mean with risk adjustment.

**Construct validity.** Dehydration is related to most other ACSC indicators.

***Discussion***

Dehydration can for the most part be treated in an outpatient setting, but is potentially fatal for elderly, very young children, frail patients, or patients with serious comorbid conditions. Little evidence exists regarding the validity of this indicator, and as such there is little guidance as to potential causes of admissions. Areas may wish to examine the outpatient care for dehydration, to identify potential processes of care that may reduce admission rates.

Admission for dehydration is somewhat common, suggesting that the indicator will be measured with adequate precision. Our empirical results confirmed that this indicator is measured with adequate precision for use as a quality indicator. The high signal ratio suggests that the observed variation is likely to reflect true differences in performance.

This indicator is subject to some minimal bias. Risk adjustment appears to affect the areas with the highest and lowest rates modestly. Age may be a particularly important factor, and should be risk adjusted for. It is unknown how other clinical factors would impact this measure. Some dehydration care takes place in an emergency room setting. As such, considering inpatient and emergency room data together may give a more accurate picture of this indicator.

Dehydration is an avoidable hospitalization/ ambulatory care sensitive condition indicator. These indicators are not measures of hospital quality, but rather measures of outpatient and other healthcare. These measures would be of most interest to comprehensive health care delivery systems, such as some health maintenance organizations, or public health officials. ACSC indicators are correlated with each other and may be used in conjunction as an overall examination of outpatient care.

Areas may wish to identify hospitals that contribute the most to the overall area rate for this indicator. The patient populations served by these hospitals may be a starting point for interventions.

Overall, this indicator is recommended for inclusion in the HCUP II QI set, though it is recommended that it be used in conjunction with other ACSC indicators. It received an empirical rating of 14 out of 26. This indicator is recommended with several caveats of use. As an ACSC indicator, this indicator may be viewed as a proxy for actual quality problems. This indicator has unclear construct validity, as this indicator has not been validated except as part of a set of indicators. Further, it is possible that providers may reduce admission rates without actually improving quality, by shifting care to an outpatient setting. Caution should be maintained for admission rates that are drastically below or above the average or recommended rates.

## INDICATOR 18: ACSC: BACTERIAL PNEUMONIA ADMISSION RATE

| | |
|---|---|
| Indicator | Area level admission rate for bacterial pneumonia. |
| Relationship to Quality | Proper outpatient treatment may reduce admissions for bacterial pneumonia in non-susceptible individuals. As such lower rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Admissions for bacterial pneumonia per 100,000 population. |
| Outcome of Interest | Discharges with ICD-9 principal diagnosis codes for bacterial pneumonia.<br><br>Age less than 65 years.*<br><br>Exclude patients with sickle cell anemia or HB-S disease disease (see Appendix 6) in any field.<br>Exclude patients <2 months (8 weeks) of age.<br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |
| Population at Risk | Population in MSA or county, age less than 65 years.* |

* Rate can also be calculated for age 65 and older.

***Evidence from the literature***

**Face validity:** Bacterial pneumonia is a relatively common acute condition, treatable for the most part with antibiotics. If left untreated in a susceptible individual, pneumonia can lead to death (see pneumonia in-hospital mortality indicator). The elderly are particularly susceptible to pneumonia. A vaccine has been developed, which is used primarily in this population. This vaccine has been shown to be 45% effective in preventing hospitalizations in the elderly during peak seasons.[507] A 1995 survey of Americans older than 65 years found that only 35.6% of respondents reported ever receiving the vaccine. The same study found that minority populations reported a lower rate of vaccination (Hispanic, 24%, Black, 20%) as compared to Whites (37%). Overall state rates were not associated with minority rates of vaccination. This study is limited by the use of self-report, and therefore, vaccination may be underreported.[508]

Appropriateness of admissions for bacterial pneumonia may account for some variation in admission rates. One study of emergency department triage strategies found a relatively modest impact on hospitalization rates for low risk cases – cases which may have been treatable on an outpatient basis.[509]

**Precision:** Pneumonia is a very common cause of hospital admission, particularly in the elderly, suggesting that relatively precise estimates of area rates should be feasible. However, little evidence exists on the precision or variation in pneumonia admission rates.

**Minimum bias:** We found no literature on the potential bias of this indicator. As the elderly are more susceptible to pneumonia, rates may be associated with the age structure of the population. Rates may also differ with comorbid diseases and socioeconomic status. Immunosuppressed patients are more likely to both develop and require hospitalization as a result of pneumonia. Some causes of immunosupression may vary systematically by area. However, we found no evidence that comorbidities or other risk factors that may vary systematically by area significantly affect the incidence of hospitalization for pneumonia. Physician thresholds for admitting patients with pneumonia also differ, which may contribute to observed differences in admission rates.

**Construct validity:** We found little literature on admission for pneumonia as an ambulatory care sensitive condition indicator. Pneumonia was originally included in Weissman's[276] set of avoidable hospitalization indicators. This set was developed by physician panels. Evidence on sets of ambulatory care sensitive condition indicators are summarized at the beginning of this section, and should be referred to for this indicator.

Two studies of ACSC conditions have examined pneumonia independently. Millman et al.[491] reported that low-income zip codes had 5.4 times more pneumonia hospitalizations per capita than high-income zip codes in 11 states in 1988. Billings et al.[489] found that low-income zip codes in New York City (where at least 60% of households earned less than $15,000 in 1988, based on adjusted 1980 Census data) had 5.4 times more pneumonia hospitalizations per capita than high-income zip codes (where less than 17.5% of households earned less than $15,000). Household income explained 53% of the variation in pneumonia hospitalization rates at the zip code level.

**Fosters true quality improvement:** We found no evidence of the impact such a measure would have on quality. As some cases of pneumonia can be managed on an outpatient basis, it is possible that use of this indicator may encourage a shift to outpatient care without actually affecting patient outcomes. Such a shift might be inappropriate for more severely ill patients.

**Prior use:** This indicator was included in Weissman's set of avoidable hospitalizations.[276] Immunization preventable pneumonia was included in the HCUP I indicator set.

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw area level rate/standard deviation | 395.6, 208.5 | |
| Systematic area-level standard deviation* | 0.09% | High |
| Area variation as a percentage of total variation* | 0.03% | High |
| Signal ratio* | 92.9% | Very High |
| R-Square* | 93.1% | Very High |
| * age- and gender- adjusted | | |
| *Minimum Bias – age-sex risk adjustment* | | |
| Signal variance change with risk adjustment | No change | Good |
| Absolute impact: | | |
| Average absolute change (in %) | 11.1% | Good |
| Relative impact: | | |
| Rank correlation | 0.923 | Good |
| Percent remaining in high decile/low decile | 68.2% / 90.9% | Good |
| Percent changing more than 2 deciles | 9.7% | Good |

**Precision.** This indicator is precise, with a raw area level rate of 395.6 per 100,000 population and a standard deviation of 208.5. The systematic area level standard deviation is high, at 0.09%. The area level variation also accounts for a high percentage of total variation, at 0.03%. The signal ratio is very high, at 92.9%. This means that it is likely that the observed differences in area performance represent true differences in area performance. The very high R-square reflects the large proportion of signal that is extractable using multivariate methods. However, multivariate techniques have little additional impact, due primarily to the already large signal ratio.

**Bias.** Signal variance does not change significantly with risk adjustment. The indicator performs well on multiple measures of minimum bias. The rank correlation is good at 0.923. Risk adjustment appears to affect the highest decile disproportionately to the lowest decile, as 68.2% of areas in the highest decile and 90.9% of the lowest decile remain after risk adjustment. However, this is consistent with other indicators. The indicator still performs well relative to other indicators on this measure. The absolute magnitude of the impact is moderate, as is the relative impact.

**Construct validity.** Bacterial pneumonia is related to most other ACSC indicators.

***Discussion***

Bacterial pneumonia can for the most part be treated using available antibiotics, however, low quality treatment may increase the admission rate for this condition. The elderly population is particularly susceptible to pneumonia, and in this population a vaccine is suggested to prevent pneumonia. Little evidence exists regarding the validity of this indicator, and as such there is little guidance as to potential causes of admissions. Areas may wish to examine the outpatient care for pneumonia and pneumococcal vaccination rates, to identify potential processes of care that may reduce admission rates.

Appropriateness of admissions appears to be a particular problem for this indicator. High rates may reflect a large number of inappropriate admissions, and/or poor quality outpatient care, among other things. While some view inappropriate admissions as a quality concern, others do not, as it concerns mainly resource overutilization and may not pose a significant additional risk to the patient.

Admission for bacterial pneumonia is relatively common, suggesting that the indicator will be measured with good precision. Our empirical results showed that this indicator is measured with adequate precision for use as a quality indicator. The very high signal ratio suggests that the observed variation is likely to reflect true differences in performance. Multivariate techniques do not appear to have substantial additional impact. Thus, either univariate or multivariate smoothing is recommended.

This indicator is subject to some moderate bias. Relative and absolute performance change somewhat with age-sex risk adjustment. In particular risk adjustment appears to affect the areas with the highest rates the most. Age may be a particularly important factor, and should be risk adjusted for. It is unknown how other clinical factors would impact this measure. In addition, some pneumonia care takes place in an emergency room setting. As such, considering inpatient and emergency room data together may give a more accurate picture of this indicator.

Bacterial pneumonia is an avoidable hospitalization/ ambulatory care sensitive condition indicator. These indicators are not measures of hospital quality, but rather measures of outpatient and other healthcare. These measures would be of most interest to comprehensive health care delivery systems, such as some health maintenance organizations, or public health officials.

ACSC indicators are correlated with each other and may be used in conjunction as an overall examination of outpatient care.

Areas may wish to identify hospitals that contribute the most to the overall area rate for this indicator. The patient populations served by these hospitals may be a starting point for interventions.

Overall, this indicator is recommended for inclusion in the HCUP II QI set, though it is recommended that it be used in conjunction with other ACSC indicators. It received an empirical rating of 17 out of 26. This indicator is recommended with several caveats of use. As an ACSC indicator, this indicator may be viewed as a proxy for actual quality problems. This indicator has unclear construct validity, as this indicator has not been validated except as part of a set of indicators. Further, it is possible that providers may reduce admission rates without actually improving quality, by shifting care to an outpatient setting. Caution should be maintained for admission rates that are drastically below or above the average or recommended rates.

Section 3.E. Indicator 18. Bacterial Pneumonia Admission Rate.

## INDICATOR 19: ACSC: URINARY INFECTION ADMISSION RATE

| | |
|---|---|
| Indicator | Area level admission rate for urinary infection. |
| Relationship to Quality | Proper outpatient treatment may reduce admissions for urinary infection. As such lower rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Admissions for urinary infection per 100,000 population. |
| Outcome of Interest | Discharges with ICD-9 principal diagnosis code of urinary tract infection per 100,000 population (see Appendix 6).<br><br>Age less than 65 years.*<br><br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |
| Population at Risk | Population in MSA or county, age less than 65.* |

* Rate can also be calculated for age 65 and older.

***Evidence from the literature***

**Face validity:** Urinary infection (UTI) is a common acute condition. Uncomplicated urinary tract infections are treatable with antibiotics on an outpatient basis. If left untreated or incompletely treated in a susceptible individual, urinary tract infections can spread into the kidneys (pyelonephritis) or develop into septicemia. Among children, admission for UTI is associated with physiological abnormalities and is rare.

**Precision:** We found little evidence on the precision of this indicator. One study did note that UTIs and kidney infections accounted for 10.6% of total admissions for ACSCs.[282]

**Minimum bias:** Urinary tract infections are a somewhat common cause of hospitalization. We found no literature on the potential bias of this indicator. We found no evidence that comorbidities or other risk factors that may vary systematically by area may increase the incidence of hospitalization for UTI. Thresholds for admission of patients with urinary tract infections may differ across areas, potentially contributing to observed differences in area hospitalization rates.

**Construct validity:** We found little literature on admission for urinary infection as an ambulatory care sensitive condition indicator. UTI was originally included in both John Billings'[489] set of indicators developed for the United Hospital Fund of New York, and in Weissman's[276] set of indicators. These sets were developed by physician panels. Evidence on sets of ambulatory care sensitive condition indicators are summarized at the beginning of this section, and should be referred to for this indicator.

Two studies of ACSC indicators reported validation work for UTI independent of measure sets. Millman et al.[491] reported that low-income zip codes had 2.8 times more UTI

hospitalizations per capita than high-income zip codes in 11 states in 1988. Billings et al.[489] found that low-income zip codes in New York City (where at least 60% of households earned less than $15,000 in 1988, based on adjusted 1980 Census data) had 2.2 times more UTI hospitalizations per capita than high-income zip codes (where less than 17.5% of households earned less than $15,000). Household income explained 28% of the variation in UTI hospitalization rates at the zip code level.

**Fosters true quality improvement:** We found no evidence of the impact such a measure would have on quality. As most UTIs can be managed on an outpatient basis, it is possible that a shift to outpatient care may occur. The "appropriate" rate of inpatient treatment for UTI is unknown, and thus there is little evidence on whether the shift to outpatient care would reduce quality of care.

**Prior use:** This indicator was included in Billings[489] set of Ambulatory Care Sensitive Conditions, developed in conjunction with the United Hospital Fund of New York, and in Weissman's set of avoidable hospitalizations.[276]

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw area level rate/standard deviation | 145.1, 89.5 | |
| Systematic area-level standard deviation* | 0.04% | Moderate |
| Area variation as a percentage of total variation* | 0.01% | Moderate |
| Signal ratio* | 84.9% | High |
| R-Square* | 85.4% | High |
| * age- and gender- adjusted | | |
| *Minimum Bias – age-sex risk adjustment* | | |
| Signal variance change with risk adjustment | No change | Good |
| Absolute impact: | | |
| Average absolute change (in %) | 11.5% | Good |
| Relative impact: | | |
| Rank correlation | 0.914 | Good |
| Percent remaining in high decile/low decile | 68.2% / 90.9% | Good |
| Percent changing more than 2 deciles | 8.8% | Good |

**Precision.** This indicator is precise, with a raw area level rate of 145.1 per 100,000 population and a standard deviation of 89.5. The systematic area level standard deviation is moderate, at 0.04%. The area level variation also accounts for a high percentage of total variation, at 0.01%. This means that relative to other indicators a higher percentage of the total variation reflects unobserved patient differences. The signal ratio is high, at 84.9%. This means that it is likely that the observed differences in area performance represent true differences in area performance, though this is lower than other indicators. The high R-square reflects the large proportion of signal variance that can be extracted using multivariate techniques, though lower than some other indicators. Multivariate techniques have little additional impact.

**Bias.** Signal variance does not change with risk adjustment. The indicator performs well on multiple measures of minimum bias. The rank correlation is good at 0.914. Risk adjustment appears to impact the highest decile disproportionately to the lowest decile, as 68.2% of areas in the highest decile and 90.9% of the lowest decile remain after risk adjustment. The absolute magnitude of the impact is moderate, as is the relative impact.

**Construct validity.** Urinary infection is related to most other ACSC indicators.

***Discussion***

Urinary tract infection can for the most part be treated in an outpatient setting, but may progress to more clinically significant infections, such as pyelonephritis, in vulnerable individuals with inadequate treatment. Little evidence exists regarding the validity of this indicator, and as such there is little guidance as to potential causes of admissions. Areas may wish to examine the outpatient care for urinary tract infection, to identify potential processes of care that may reduce admission rates.

Admission for urinary tract infection is uncommon, suggesting that there may be potential problems with precision. Our empirical results demonstrates that this indicator is measured with adequate precision for use as a quality indicator. The high signal ratio suggests that observed variation is likely to reflect true differences in performance. Multivariate techniques do not appear to have substantial additional impact. Therefore, univariate or multivariate smoothing is recommended for this indicator.

This indicator is subject to some moderate bias. Relative and absolute performance change somewhat with age-sex risk adjustment, and so we recommend adjusting for age and sex. Risk adjustment appears to affect the areas with the highest rates the most, though this impact is moderate in comparison to other indicators. Using this indicator without risk-adjustment may result in the misidentification of some areas as outliers. It is unknown how other clinical factors would impact this measure. In addition, some urinary tract infection care takes place in an emergency room setting. As such, considering inpatient and emergency room data together may give a more accurate picture of this indicator.

Urinary tract infection is an avoidable hospitalization/ ambulatory care sensitive condition indicator. These indicators are not measures of hospital quality, but rather measures of outpatient and other healthcare. These measures would be of most interest to comprehensive health care delivery systems, such as some health maintenance organizations, or public health officials. ACSC indicators are correlated with each other and may be used in conjunction as an overall examination of outpatient care.

Areas may wish to identify hospitals that contribute the most to the overall area rate for this indicator. The patient populations served by these hospitals may be a starting point for interventions.

Overall, this indicator is recommended for inclusion in the HCUP II QI set, though it is recommended that it be used in conjunction with other ACSC indicators. It received an empirical rating of 11 out of 26. This indicator is recommended with several caveats of use. As an ACSC indicator, this indicator may be viewed as a proxy for actual quality problems. This indicator has unclear construct validity, as this indicator has not been validated except as part of a set of indicators. Further, it is possible that providers may reduce admission rates without actually improving quality, by shifting care to an outpatient setting. Caution should be maintained for admission rates that are drastically below or above the average or recommended rates.

## INDICATOR 20: ACSC: PERFORATED APPENDIX ADMISSION RATE

| | |
|---|---|
| Indicator | Area level admission rate for perforated appendix. |
| Relationship to Quality | Timely diagnosis and treatment may reduce the incidence of perforated appendix. As such lower rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Admissions for perforated appendix per 100 admissions for perforated appendix within MSA or county. |
| Outcome of Interest | Discharges with ICD-9 diagnosis code for perforations or abscesses of appendix in any field per 100 discharges with diagnosis code for appendicitis within area (see Appendix 6).<br><br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |
| Population at Risk | Discharges with diagnosis code for appendicitis within MSA or county. |

***Evidence from the literature***

**Face validity:** Perforated appendix may occur when appropriate treatment for acute appendicitis is delayed for any variety of reasons. Such delay could result from access to care problems, patients failing to interpret symptoms as potentially important, misdiagnosis and other delays in obtaining surgery.

**Precision:** Perforated appendix is a relatively common condition, occurring in ¼ to 1/3 of hospitalized acute appendicitis patients.[510] Thus, it is likely to be measured with good precision.

**Minimum bias:** Observational studies utilizing large administrative databases have noted higher rates of perforated appendix in males,[510] patients with mental illness or substance abuse disorders,[510] diabetics,[510] blacks,[510] and children under the age of four (though appendicitis is rare in this age group).[511] If areas have an unusually high proportion of patients with diabetic comorbidity, or patient with substance abuse or psychiatric comorbidity may have higher rates of perforated appendix that may not be due to actual differences in quality of care. However, since appendicitis is relatively rare as compared with these comorbidities, it is unlikely that an area rate would change significantly due to perforated appendix in these populations.

**Construct validity:** Three recent studies have examined perforated appendix as a measure of access to care. The first examined all California adult admissions (18-64 years of age), with acute appendicitis (96,587 cases).[510] They found that patients with either no insurance or Medicaid had just under 50% greater risk of perforated appendix than HMO covered patients. They interpreted this result to be evidence of potential access to care problems. They also found a 20% increased risk in patients with private fee-for-service insurance. In a follow-up to this

study, Blumberg et al. noted that when examining the high rate of perforated appendix in the black population at Kaiser, patients with almost identically comprehensive coverage, black patients did not have higher rates of delayed admissions, a possible indication of poor quality care. They postulated that rather, delay in seeking care may explain some of the differences observed.[512] Thus, it is unclear whether the higher rate of perforated appendix is due to quality of care problems, perceived access to care difficulties, or other reasons for the delay in seeking care.

The second study took a similar approach in analyzing pediatric admissions for acute appendicitis in Washington state.[511] The rate of perforated appendix was again increased in Medicaid patients [Adj. OR 1.3 ,95% CI (1.2-1.4)]. Another study in a pediatric population examined reasons for delay to surgery and insurance status in a New York pediatric population through retrospective chart review. They noted that Medicaid or uninsured children had both a higher perforation rate and a longer duration of symptoms before presenting to a health care professional as compared to HMO or private fee for service insured children. There were no differences between the types of insurance in the time to surgery after presentation.[513] Unfortunately the authors did not analyze how much of the variance in perforated appendix could be explained by delays in seeking care.

Weissman et al. their analysis of avoidable hospitalizations found that uninsured had a relative risk of 1.14-1.20 of admission for ruptured appendix after adjusting for age and sex. Medicaid patients had a relative risk of .45-.58, suggesting that in at least this case, Medicaid patients are not at increased risk for ruptured appendix.[276]

**Fosters true quality improvement:** Rates of perforated appendix could be increased by increasing the denominator, total patients admitted for appendicitis, by increasing the number of patients with borderline or questionable cases undergoing appendectomy.[510]

**Prior use.** Perforated appendix was included in both the previous HCUP I indicator set, as well as in Weissman's set of avoidable hospitalizations.[276]

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw area level rate/standard deviation | 33.3%, 14.4% | |
| Systematic area-level standard deviation* | 2.75% | Very High |
| Area variation as a percentage of total variation* | 0.34% | Very High |
| Signal ratio* | 26.5% | Low |
| R-Square* | 39.4% | Low |
| * age- and gender- adjusted | | |
| *Minimum Bias – age-sex risk adjustment* | | |
| Signal variance change with risk adjustment | Decreases | Fair |
| Absolute impact: | | |
| Average absolute change (in %) | 2.0% | Very Good |
| Relative impact: | | |
| Rank correlation | 0.969 | Very Good |
| Percent remaining in high decile/low decile | 90.0% / 85.0% | V.G/ Good |
| Percent changing more than 2 deciles | 2.5% | Very Good |

**Precision.** This indicator is precise, with a raw area level rate of 33.3% and a substantial standard deviation of 14.4%. The systematic area level standard deviation is very high, at 2.75%. The area level variation also accounts for a very high percentage of total variation, at

0.34%. This means that relative to other indicators, a higher percentage of the variation occurs at the area level, rather than the discharge level. However, the signal ratio is low, at 26.5%. This means that it is likely that much of the observed differences in area performance do not reflect true differences in performance. The R-Square does demonstrate the improvement in the amount of signal that can be extracted using multivariate techniques. Nonetheless, the R-square is still low at 39.4%.

**Bias.** Signal variance decreases by over 15% with risk adjustment by age and sex. This indicates that some of the apparent signal is due to systematic differences in patient demographics. The demographic (age-sex) adjustment had minimal impact on most measures of minimum bias. The rank correlation was high at 0.969. The number of areas changing more than two deciles was relatively low, at 2.5%. Relative to other indicators, this risk adjustment had a moderate impact on the lowest end of the distribution, with 85% of providers in the lowest decile remaining after risk adjustment.

**Construct validity.** Perforated appendix is negatively related to the other ACSC conditions. It is positively related to low birth weight.

***Discussion***

With prompt and appropriate care, acute appendicitis should not progress to perforation or rupture. However, rates of perforated appendix are higher in the uninsured or underinsured in both the adult and pediatric population. It is unclear whether this arises from patients failing to seek appropriate care, access to care difficulties, or misdiagnoses and poor quality care. Areas with high rates may want to investigate the reasons for delay in receiving surgery in order to target points of intervention. This may be accomplished through mechanisms such as chart reviews and other supplemental data.

In our empirical analyses we observed very high systematic variation in this condition, though the signal ratio was low, suggesting that some of the variance observed does not reflect true differences in provider performance. This indicator is measured with sufficient precision for use as a quality indicator. However, as multivariate smoothing appeared to increase the extractable signal for this indicator, it is highly recommended.

This indicator performed well on our measures of minimum bias, adjusting for age and sex. We found no evidence in our literature review that clinical characteristics that would vary systematically increase the likelihood of perforated appendix. Thus, this is unlikely to be a clinically biased indicator. Perforated appendix rates vary systematically by race, but whether this is due to poor quality care or other factors is not known.

Perforated appendix is an avoidable hospitalization/ ambulatory care sensitive condition indicator. These indicators are not measures of hospital quality, but rather measures of access to care, outpatient and other health care, and as such are defined with area level denominators. These measures would be of most interest to comprehensive health care delivery systems, such as some health maintenance organizations, or public health officials. ACSC indicators are correlated with each other and may be used in conjunction as an overall examination of outpatient care.

Areas may wish to identify hospitals that contribute the most to the overall area rate. The patient populations served by these hospitals may be a starting point for interventions. Hospital contributions to the overall area rate may be of particular utility for this indicator. As perforated appendix can arise from delay in receiving surgeries, misdiagnoses and other causes of delay in receiving surgery in emergencies rooms may contribute to the rate substantially.

Overall, this indicator is recommended for inclusion in the HCUP II QI set. It received an empirical rating of 17 out of 26. Smoothing is recommended for this indicator. This indicator is recommended with one caveat of use. As an ACSC indicator, this indicator may be viewed as a proxy for actual quality problems.

Section 3.E. Indicator 20. Perforated Appendix Rate.

## INDICATOR 21: ACSC: ANGINA WITHOUT PROCEDURE ADMISSION RATE

| | |
|---|---|
| Indicator | Area level admission rate for angina (without procedure). |
| Relationship to Quality | Proper outpatient treatment may reduce admissions for angina (excluding admission for procedures). As such lower rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Admissions for angina (without procedures) per 100,000 population. |
| Outcome of Interest | Discharges with ICD-9 principal diagnosis codes for angina (see Appendix 6).<br><br>Age 18 years to 64 years old.*<br><br>Exclude discharges with a surgical procedure in any field (01.0-86.99).<br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |
| Population at Risk | Population in MSA or county, age 18-64 years.* |

* Rate can also be calculated for age 65 and older.

***Evidence from the literature***

**Face validity:** Both stable and unstable angina are symptoms of potential coronary artery disease. Stable angina can be managed in an outpatient setting, using drugs such as aspirin, beta blockers, and advise to change diet and exercise habits.[514] Unstable angina may be managed on an outpatient basis, but admission is required for more severe cases, such as patients with recurrent or progressive symptoms.[515] Effective management of coronary disease reduces the occurrence of major cardiac events such as heart attacks, and may also reduce admission rates for angina.

**Precision:** Unstable angina is a very common reason for hospital admission. One study noted that angina accounted for 16.3% of total admissions for ACSCs.[282] Thus, reasonably precise estimates of area angina rates should be feasible.

**Minimum bias:** We found no literature on the potential bias of this indicator. The incidence of angina is related to the incidence of coronary artery disease (CAD), which is in turn related to the age structure and risk factors (smoking, hyperlipidemia, hypertension, diabetes), in a population. Some areas may systematically vary in the incidence of angina, as a result of these factors and potentially correlated differences in socioeconomic status. Elderly age (over 70), diabetes and hypertension have also been associated with higher risk angina.[515] Finally, physicians may differ in their thresholds for admitting patients with less stable angina. Little evidence exists on the extent to which these factors may account for area differences in angina admission rates.

**Construct validity:** We found little literature on admission for angina as an ambulatory care sensitive condition indicator. Angina was originally included in John Billings'[275] set of indicators developed for the United Hospital Fund of New York. This set was developed by a physician panel. Evidence on sets of ambulatory care sensitive condition indicators are summarized at the beginning of this section, and should be referred to for this indicator.

Two studies of ACSC indicators reported validation work for angina independent of measure sets. Millman et al.[491] reported that low-income zip codes had 2.7 times more angina hospitalizations per capita than high-income zip codes in 11 states in 1988. Billings et al.[489] found that low-income zip codes in New York City (where at least 60% of households earned less than $15,000 in 1988, based on adjusted 1980 Census data) had 2.3 times more angina hospitalizations per capita than high-income zip codes (where less than 17.5% of households earned less than $15,000). Household income explained 13% of the variation in angina hospitalization rates at the zip code level.

**Fosters true quality improvement:** We found no evidence of the impact such a measure would have on quality. As some angina can be managed on an outpatient basis, it is possible that a shift to outpatient care may occur without true changes in the occurrence of angina complications. However, it seems unlikely that severe angina, requiring observation, would be shifted to outpatient settings.

**Prior use:** This indicator was included in Billings set of Ambulatory Care Sensitive Conditions, developed in conjunction with the United Hospital Fund of New York.[275]

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw area level rate/standard deviation | 166.0, 135.7 | |
| Systematic area-level standard deviation* | 0.06% | High |
| Area variation as a percentage of total variation* | 0.04% | High |
| Signal ratio* | 91.6% | Very High |
| R-Square* | 91.9% | Very High |
| * age- and gender- adjusted | | |
| *Minimum Bias – age-sex risk adjustment* | | |
| Signal variance change with risk adjustment | No change | Good |
| Absolute impact: | | |
| Average absolute change (in %) | 10.6% | Good |
| Relative impact: | | |
| Rank correlation | 0.968 | Very Good |
| Percent remaining in high decile/low decile | 63.6% / 90.9% | Good |
| Percent changing more than 2 deciles | 3.2% | Very Good |

**Precision.** This indicator is precise, with a raw area level rate of 166 per 100,000 population and a standard deviation of 135.7. The systematic area level standard deviation is high, at 0.06%. The area level variation also accounts for a high percentage of total variation, at 0.04%. The signal ratio is very high, at 91.6%. This means that it is likely that the observed differences in area performance represent true differences in area performance. The very high R-square reflects the large proportion of signal that can be extracted using multivariate techniques. Such techniques do not have substantial additional impact, primarily due to the already very high signal ratio.

**Bias.** Signal variance does not change with risk adjustment. The indicator performs well on multiple measures of minimum bias. The rank correlation is very good at 0.968. Risk adjustment appears to disproportionately impact the high decile compared to the low decile, as 63.6% of areas in the highest decile and 90.9% of the lowest decile remain after risk adjustment. Few providers change more than two deciles. The absolute magnitude of the impact is moderate, with an average change with risk adjustment (relative to the mean) of 10.6%.

**Construct validity.** Angina is related to most other ACSC indicators.

***Discussion***

Angina without procedure is a common reason for admission. Some angina admissions may be avoidable through proper outpatient care, aimed at treating the effects of coronary artery disease. Little evidence exists regarding the validity of this indicator, and as such there is little guidance as to potential causes of admissions. Areas may wish to examine the outpatient care for angina, as well as emergency room care, to identify potential processes of care that may reduce admission rates.

Admission for angina is relatively common, suggesting that the indicator will be measured with good precision. Our empirical results showed that this indicator is measured with adequate precision for use as a quality indicator. The very high signal ratio suggests that the observed variation is likely to reflect true differences in performance. Multivariate techniques do not have substantial additional impact, and as such either univariate or multivariate techniques is recommended.

Risk adjustment does not impact this measure substantially, though there is some moderate impact. Age and sex may be particularly important factors, and should be risk adjusted for. It is unknown how other clinical factors would impact this measure. Some angina care takes place in an emergency room setting. As such, considering inpatient and emergency room data together may give a more accurate picture of this indicator.

Angina is an avoidable hospitalization/ ambulatory care sensitive condition indicator. These indicators are not measures of hospital quality, but rather measures of outpatient and other healthcare. These measures would be of most interest to comprehensive health care delivery systems, such as some health maintenance organizations, or public health officials. ACSC indicators are correlated with each other and may be used in conjunction as an overall examination of outpatient care.

Areas may wish to identify hospitals that contribute the most to the overall area rate for this indicator. The patient populations served by these hospitals may be a starting point for interventions.

Overall, this indicator is recommended for inclusion in the HCUP II QI set, though it is recommended that it be used in conjunction with other ACSC indicators. It received an empirical rating of 19 out of 26. This indicator is recommended with several caveats of use. As an ACSC indicator, this indicator may be viewed as a proxy for actual quality problems. This indicator has unclear construct validity, as this indicator has not been validated except as part of a set of indicators. Further, it is possible that providers may reduce admission rates without actually improving quality, by shifting care to an outpatient setting. Caution should be maintained for admission rates that are drastically below or above the average or recommended rates.

## INDICATOR 22: ACSC: ADULT ASTHMA ADMISSION RATE

| | |
|---|---|
| Indicator | Area level admission rate for adult asthma. |
| Relationship to Quality | Appropriate and continued treatment may reduce the incidence of exacerbation of asthma requiring hospitalization. As such lower rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Admissions for adult asthma per 100,000 population. |
| Outcome of Interest | Discharges with ICD-9 principal diagnosis codes for asthma (see Appendix 6).<br><br>Age 18-64 years.*<br><br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |
| Population at Risk | Population in MSA or county, age 18-64 years and older.* |

* Rate can also be calculated for age 65 and older.

***Evidence from the literature***

**Face validity.** Asthma is one of the most common chronic diseases in the United States, affecting nearly 15 million individuals.[516] In 1997, there were about 433,000[517] to 484,000 hospitalizations for asthma in the US,[518] of which at least 252,000 involved persons 18 or more years of age. In 1996, asthma was the 10th most common principal diagnosis in emergency department (ED) visits and the ninth most common diagnosis in hospital outpatient departments.[519] In this discussion, we will consider only admission for asthma in adults. A separate discussion of the evidence for the pediatric asthma indicator (indicator 31) is included elsewhere in this report.

There is widespread consensus, embodied in the National Asthma Education Program,[520] that asthma is a readily treatable chronic disease that can be managed effectively in the outpatient setting. As Healthy People 2010 indicates, "effective management of asthma includes four components: avoiding or controlling the factors that may make asthma worse…, taking appropriate medications tailored to the severity of the disease, objective monitoring of the disease by the patient and the health care professional, and actively involving the patient in managing the disease."

Many asthma exacerbations are preventable using inhaled anti-inflammatories such as corticosteriods and mast cell stabilizers, or treatable in the outpatient setting using beta agonists and systemic corticosteroids. Observational studies offer limited evidence that inhaled steriods may decrease risk of admission by up to 50%.[521, 522] Potential confounding factors such as asthma severity, however, limit the conclusiveness of these results. In addition, factors that

explain variation in the risk of hospitalization at the patient level may not explain variation in area-level hospitalization rates.

**Precision.** Asthma is a common cause of admission for adults, and as such this measure is likely to have adequate precision. Blustein et al. noted that asthma accounted for 5% of ACSC admissions for Medicare beneficiaries.[282] One United Kingdom study noted that 51% of health authorities changed over 10 or more places when ranked by asthma admission rates over a 1 year period (1993-1994).[522] The generalizability of this study to the United States is unknown.

**Minimum bias.** Numerous environmental risk factors for asthma have been identified, and some of these factors are more prevalent in certain communities than in others. Indoor allergens such as cockroaches and dust mites may be more common in lower-income areas, and are probably associated with increased frequency and severity of asthma symptoms.[523] Tobacco smoke is the most important indoor irritant and is a major precipitant of asthma symptoms in both children and adults.[524-527] Jindal and colleagues[526] found that exposure of adults to environmental tobacco smoke is associated with decreased pulmonary function, increased medication requirements, and more frequent absences from work. Outdoor air pollution, especially respirable particulates, may also play a role.[520, 528-532] In addition, ozone and $SO_2$ have been associated with increased emergency department visits and hospitalizations rates.[531-538] Increasing air pollution has been specifically correlated with higher admission rates in London (Ozone, $NO_2$, $SO_2$, and black smoke),[539] and Seattle (ambient air pollution)[540].

Occupation is a potentially important source of bias in comparing asthma hospitalization rates among adults. A recent review of 43 studies from 19 different countries found that at least 9%, and perhaps as much as 15% (based on the highest quality studies), of the population burden of asthma is attributable to occupational factors.[541] About 250 agents capable of causing occupational asthma have been identified .[542] These agents affect individuals in specific occupations, such as carpenters and woodworkers,[543] who may cluster in certain geographic areas.

Race represents one of the most complex potentially biasing factors for this indicator. Black patients have consistently been shown to have higher asthma admission rates,[544-546] even when stratifying for income and age.[547] One study examining differences in asthma health care utilization noted that African Americans made fewer asthma-related primary care and specialist visits than Caucasian patients (47.6% vs. 70.2% and 27% vs. 38.8%). There were no differences in hospitalization rates by race, but African-American patients had lower household incomes, and made more emergency department visits (proxy for either access to care or asthma severity).[548] Similarly, Hispanics have been shown to have higher admission rates than non-Hispanic whites (or areas with higher percentages of Hispanics have been shown to have higher admission rates), although none of these studies controls for SES.[545, 549, 550] To the extent that true differences in disease prevalence or severity are responsible for racial variation in hospitalization rates, race should be adjusted for in comparing asthma hospitalization rates across areas. On the other hand, to the extent that minority patients have less access to care or poorer quality of outpatient care, race should not be adjusted for.

**Construct validity.** Little evidence has been reported conclusively attaching poor quality of care to higher area admission rates. However, numerous studies have shown that asthma hospitalization rates are associated with socioeconomic factors, including median household income (at the area level) and lack of insurance (at the individual level). A study of asthma hospitalization rates in California in 1993 (ages 0-64) found that areas with median household incomes under $35,000 had hospitalization rates that were 1.5 times higher than areas with

higher median incomes.[547] In Boston, in 1992, age and gender standardized hospitalization rates (all ages) were correlated with percentage poverty in an area (r=0.68), percentage holding a bachelor's degree (r=-0.61), and income (r=-0.51).[550] Within New York City in 1994, asthma hospitalization rates were negatively correlated with a zip code area's median household income (r=-0.67), and positively correlated with the percentage of minorities in the population (r=0.82).[549] These findings confirm an earlier study by Billings et al.,[489] who reported 6.4-fold variation in asthma hospitalization rates at the zip code level in New York City in 1988, with 70% of this variation explainable by the percentage of households with annual income below $15,000. Millman et al.[491] reported that low-income zip codes had 5.8 times more asthma hospitalizations per capita than high-income zip codes in 11 states in 1988. Using New York State data, Lin et al showed that hospitalization rates were higher in areas with higher poverty, unemployment, minority populations, and lower education levels.[545] Even in England, 45% of the variation in asthma hospitalization rates across 90 family health services authorities in 1990-95 was attributable to socioeconomic factors, plus the availability of secondary care.[551] To our knowledge, only one study has reported partial correlations;[552] it found that that in New York City, the percentage of African-American residents was the strongest predictor, and median household income was the next strongest predictor, of asthma hospitalization rates.

Similar findings have been demonstrated at the individual level. For example, Weissman et al.[276] noted that adjusted relative rate of adult asthma hospitalization for uninsured persons, compared with privately insured persons, was 1.42 (95% CI, 1.20-1.63) in Massachusetts and 1.19 in Maryland (95%CI, 0.92-1.45). The comparable rates for Medicaid beneficiaries were 1.84 (95% CI, 1.63-2.05) and 1.61 (95% CI, 1.33-1.90), respectively. Bierman et al.[553] noted that the asthma hospitalization rate among persons with private insurance nationwide was 33% lower, and that among Medicaid beneficiaries was significantly higher, than that in the general population.

The observation that asthma admission rates are higher in areas with low SES has led some researchers to hypothesize that lack of access to care, or poor quality outpatient care, may lead to higher admission rates. Bindman et al.[284] showed that asthma hospitalization rates across 41 sampled areas in California were significantly correlated (r=0.47) with self-rated access to needed medical care, according to community telephone surveys. Although analyses of the National Health and Nutrition Examination Survey found that Medicaid enrollment and Spanish language preference were associated with inadequate asthma therapy, these deficiencies in care were not directly linked to hospitalizations.[554] Studies from other settings have shown that African-American asthmatics tend to have fewer scheduled primary care visits, and more hospitalizations and emergency room visits, than White asthmatics.[555, 556] African-Americans' use of asthma medications may also be less consistent with current practice guidelines.[557]

Some weak evidence indicates that patients admitted to the hospital often receive suboptimal outpatient care, as measured by guideline adherence. The National Asthma Education Program (NAEP) Guidelines suggest that patients receive proper asthma education, including use of an MDI and written action plans in case of exacerbation.[520] One study of patients hospitalized in an inner-city teaching hospital noted that only 28% had received an action plan, and 11% could not demonstrate proper use of an MDI, despite reporting having been shown previously by health care personnel. In addition, 69 of the 101 patients reported were prescribed theophylline without first being prescribed anti-inflammatory inhalers. Finally, 60% of the patients who contacted physicians during the current exacerbation reported that the health care provider made no changes in treatment.[558] Similar results, with 95% of patients failing to use an action plan

and 51% having inadequate knowledge, were reported from an Australian hospital.[559] Neither of these studies had a control group with non-hospitalized asthmatics, and both relied on self-reported treatment data.

Few studies have directly linked high-quality processes of outpatient care with lower hospitalization rates at either the area or the individual level. An in-depth study of asthma treatment practices in New Haven, Boston, and Rochester found that the community with the highest asthma hospitalization rate (Boston) also had lower use of inhaled anti-inflammatory agents and oral steroids. The threshold for admission also appeared to be lower in Boston, as fewer of the admitted patients were hypoxemic, relative to the other cities.[560] One case control study from a large health maintenance organization established that not having a written asthma management plan was a strong risk factor for asthma hospitalization (after adjusting for severity of asthma), but the use of antiinflammatory medications was not.[561] Although these studies focused on children rather than adults, the results provide limited support for the construct validity of the asthma hospitalization rate as an indicator of access to high-quality outpatient care.

**Fosters true quality improvement.** We located no studies discussing the ability of this indicator to foster true quality improvement. One study from the United Kingdom[551] argues that area admission rates for asthma are not a good measure of quality, due to the confounding factors discussed above. While we located no studies examining the potential for gaming for this indicator, it is possible that patients who present to outpatient clinics or emergency rooms as candidates for admission would not be admitted, but rather treated in the outpatient or ER setting. There is little evidence currently to suggest that asthmatics are being inappropriately denied admission to the hospital, although this problem could emerge in the future.

**Prior use.** Most published indicators of asthma admission include pediatric as well as adult patients in a single indicator, though they are separate indicators in this report. Children under the age of 5 are sometimes excluded. Admission for asthma has been implicated as an ambulatory care sensitive condition, included in both the preventable hospitalization set of Weissman et al.[276] and the set of ACS conditions developed by Billings et al.,[489] in conjunction with the United Hospital Fund of New York. The UK National Health Service has designated asthma admission as a High Level Performance Indicator. In addition, Healthy People 2010 has set a goal to reduce the nationwide asthma admission rate from 12.5 (in 1998) to 7.7 per 10,000 children and adults aged 5-65 years, and from 17.7 to 11 per 10,000 adults over 65 years of age.[5]

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw area level rate/standard deviation | 107.9, 81.7 | |
| Systematic area-level standard deviation* | 0.04% | Moderate |
| Area variation as a percentage of total variation* | 0.02% | Moderate |
| Signal ratio* | 83.6% | High |
| R-Square* | 84.2% | High |
| * age- and gender- adjusted | | |
| *Minimum Bias – age-sex risk adjustment* | | |
| Signal variance change with risk adjustment | No change | Good |
| Absolute impact: | | |
| Average absolute change (in %) | 3.8% | Very Good |
| Relative impact: | | |
| Rank correlation | 0.989 | Very Good |

| | | |
|---|---|---|
| Percent remaining in high decile/low decile | 86.4% / 95.5% | Very Good |
| Percent changing more than 2 deciles | 0% | Very Good |

**Precision.** This indicator is adequately precise, with a raw area level rate of 107.9 per 100,000 population and a standard deviation of 81.7. The systematic area level standard deviation is moderate, at 0.04%. The area level variation also accounts for a moderate percentage of total variation, at 0.02%. These suggests that relative to other indicators a lower proportion of variance occurs at the area level, rather than the discharge level. The signal ratio is high, at 83.6%. This means that it is likely that the observed differences in area performance represent true differences in area performance, though some reflects unsystematic differences. The high R-square reflects the large proportion of signal that can be extracted using multivariate techniques, though lower than other indicators. Such techniques do not have substantial additional impact.

**Bias.** Signal variance does not change with risk adjustment. The indicator performs well on multiple measures of minimum bias. The rank correlation is very good at 0.989. Risk adjustment does not appear to impact the extremes of the distribution substantially. No providers change more than two relative deciles. The absolute magnitude of the impact is minimal, with an average change with risk adjustment (relative to the mean) of 3.8%.

**Construct validity.** Adult asthma is related to most other ACSC indicators.

***Discussion***

Asthma is one of the most common reasons for hospital admission and emergency room care. Nonetheless, it is widely accepted that most cases of asthma can be managed with proper ongoing therapy on an outpatient basis. Our literature found some evidence that proper use of inhaled steriods may decrease asthma exacerbation, though this evidence is weak. Many studies have associated increased asthma hospitalization rates with lower socioeconomic status, though in many of these studies confounding factors were not controlled for. Surveys for patients admitted for asthma in low income areas have found inadequate outpatient care. One well designed study noted that 70% of the variance in asthma admission rates is explainable by area self-rated access to care.

Given the large number of asthma hospitalizations, we would expect that this indicator would be adequately precise. Our empirical results confirmed that this indicator is measured with adequate precision for use as a quality indicator, though less than some other indicators. The high signal ratio suggests that most of the observed variance reflects actual differences in performance. Multivariate techniques do not have substantial additional impact, and as such either univariate or multivariate smoothing is recommended.

Our empirical tests demonstrate that this indicator is not subject to substantial bias. Risk adjustment by age and sex did not have substantial impact on the performance of areas. Nonetheless, some factors other than age and sex, may vary systematically by area and may also impact the hospitalization rate for asthma. Environmental factors, such as air pollution, occupational exposure to irritants, or other exposure to allergens (i.e. cockroach, dust mite), have been shown to increase hospitalization rates or exacerbate asthma symptoms. Areas with high rates may wish to examine these factors relative to other areas when interpreting performance. While race has been shown to be associated with differences in admission rates, it is unclear whether this is due to differences in severity of disease or inadequate access to care.

Asthma is an avoidable hospitalization/ ambulatory care sensitive condition indicator. These indicators are not measures of hospital quality, but rather measures of outpatient and other

healthcare. These measures would be of most interest to comprehensive health care delivery systems, such as some health maintenance organizations, or public health officials. ACSC indicators are correlated with each other and may be used in conjunction as an overall examination of outpatient care.

Areas may wish to identify hospitals that contribute the most to the overall area rate for this indicator. The patient populations served by these hospitals may be a starting point for interventions.

Most published studies and the Healthy People 2010 indicator combine admission rates for children and adults, though in this report they remain separate indicators. Thus, areas may wish to examine this indicator together with the pediatric asthma indicator (indicator 31).

This indicator is recommended for inclusion in the HCUP II QI set. It received an empirical score of 16 out of 26. It is recommended with several caveats of use. As an ACSC indicator, this indicator may be viewed as a proxy for actual quality problems. Further, it is possible that providers may reduce admission rates without actually improving quality, by shifting care to an outpatient setting. Caution should be maintained for admission rates that are drastically below or above the average or recommended rates.

INDICATOR 23: ACSC: CHRONIC OBSTRUCTIVE PULMONARY DISEASE (COPD) ADMISSION RATE

| | |
|---|---|
| Indicator | Area level admission rate for COPD. |
| Relationship to Quality | Proper outpatient treatment may reduce admissions for COPD. As such lower rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Admissions for COPD per 100,000 population. |
| Outcome of Interest | Discharges with ICD-9-CM principal diagnosis code for COPD in any diagnosis field (see Appendix 6)<br><br>Age 18-64 years.*<br><br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |
| Population at Risk | Population in MSA or county, age 18-64 years.* |

* Rate can also be calculated for age 65 and older.

***Evidence from the literature***

Chronic obstructive pulmonary disease (COPD) consists of three primary diseases, asthma, emphysema, and chronic bronchitis. Though each disease causes respiratory dysfunction, each has somewhat distinct etiologies, treatments, and outcomes. Since admission for asthma is considered in a separate indicator, asthma will not be discussed in this section. Only the evidence for COPD as it relates to emphysema and chronic bronchitis will be discussed.

**Face validity**. The incidence of COPD has been increasing in the last decade,[562] as have admissions for COPD. Admissions for COPD include exacerabations of COPD, respiratory failure, and rarely lung volume reduction surgery or lung transplantation. COPD accounts for over $20 billion in health care expenditures. Data suggests that more than 2/3 of these costs are incurred by only 10% of persons with COPD.[563]

Practice guidelines for COPD have been set forth and published over the past decade.[564] Based on consensus, the three major guidelines (European Respiratory Society, American Thoracic Society and British Thoracic Society) agree that appropriate care for COPD includes spirometry, medication, and monitoring. All also agree that advice to quit smoking is a critical intervention. Specific recommendations for the indications of drug use vary depending on the nature of disease, though they often include bronchodialators (combined $beta_2$-agonist and anticholinergic), theophylline and corticosteriods. These guidelines are based on consensus statements, not empiric evaluation of the evidence, and call for the need for more research on the effectiveness of treatments.[565]

With appropriate outpatient treatment and compliance, hospitalizations for the exacerbations of COPD and decline in lung function should be minimized. COPD was selected by a physician panel as an ambulatory care sensitive condition.[275]

**Precision**. COPD is a common disease that accounts for a substantial number of hospital admissions (>13% of admissions),[566] suggesting that reasonably precise area rate estimates are feasible. We were unable to identify studies that examined geographic variation in admission for COPD. Slight seasonal variations have been noted in younger age groups, though in patients over 65 no seasonal variations in hospital admissions have been noted.[567] Billings' original study from New York reported only 1.8-fold variation in COPD hospitalization rates, with a coefficient of variation of 0.742.[489]

**Minimum bias**. Exacerbations of COPD and the rate of pulmonary function decline can be affected by patient characteristics. Cigarette smoking is a leading cause of COPD. Smoking after the development of COPD has been shown to accelerate the rate of pulmonary decline, with quitters sustaining less pulmonary decline that continuous smokers.[568] One study showed a small but significant increase in pulmonary decline in patients that quit smoking then restarted, compared with patients that never quit.[569] Thus, patient compliance is an important determinant of COPD admission rates. Actual smoking cessation by COPD patients is less than optimal, with many patients failing to comply despite ambulatory interventions. Even with intensive smoking cessation interventions, continuous quit rates remain at around 20%.[568] Lower rates have been reported in other studies with different patient populations, suggesting that patient characteristics may influence responsiveness to treatment.

Other factors have been associated with increased hospitalizations for COPD. For example, the association between lower socio-economic status and increased likelihood of hospital admission [570] may be due to higher occupational exposure to harmful substances, smoking rates, and infection rates postulated to exist among lower socio-economic groups.[570] However, Billings' original study from New York reported only 3% of variance explained by household income.[489] In addition, COPD is a progressive disease, and disease severity varies considerably across patients and over time. As lung function declines with time in COPD, older patients have higher rates of pulmonary decline, but heavier smoking and genetic factors also influence the rate of decline.

One environmental factor, daily increases in air pollution, has been associated with increased daily rates of COPD hospitalization.[571-579] The robustness of the association, and the type and amount of air pollution affecting admissions, varies between study. Few studies have examined air pollution differences between geographical areas and its association with admission for COPD.

Thus, in addition to direct measures of disease severity, smoking status, age, and socio-economic factors may increase the likelihood of admission for COPD. These factors are candidates for use in risk-adjustment models.

One study did indicate that COPD in younger populations might include some miscoding for acute bronchitis.[567] It is unknown whether such coding differences lead to biases in area-level estimates of COPD admission rates.

**Construct validity**. The extent to which the admission rate for COPD relates to the quality of the outpatient health care provided has not been widely evaluated, though some studies have examined readmission rates. Weissman et al. did not evaluate COPD, but Bindman et al. reported that self-reported access to care explained 27% of the variation in COPD hospitalization rates (e.g., less than for asthma, CHF, or diabetes) at the zip code cluster level.[284] Millman et al.[491] reported that low-income zip codes had 5.8 times more COPD hospitalizations per capita than high-income zip codes in 11 states in 1988. However, Billings et al.'s[489] findings were weaker; low-income zip codes in New York City (where at least 60% of households earned less

than $15,000 in 1988, based on adjusted 1980 Census data) had 1.8 times more COPD hospitalizations per capita than high-income zip codes (where less than 17.5% of households earned less than $15,000). Household income explained only 3% of the variation in COPD hospitalization rates at the zip code level.

Some articles discuss the adherence to practice guidelines and other therapies by both physicians and patients, though they do not generally examine the relationship between adherence rates and hospitalization rates for COPD. One study found varying rates of physician compliance with practices expected to improve quality for patients admitted for COPD. In this study, over 2/3 of physicians performed a complete history and physical examination and some nutritional screening. Almost all provided some sort of appropriate pharmacologic treatment (96%). However, few gave smoking cessation advice during the hospitalization (23%), and almost none provided appropriate discharge planning and education (.2%).[580] Another study found that often physicians did not adhere to practice guidelines. In this study, smoking cessation counseling was provided in 14.3% of ambulatory visits. Physicians prescribed medication contrary to indications. More than ¼ of visits resulted in a prescription of theophylline, though this drug is considered for use as a "step 3" drug, and 5% of visits resulted in a prescription for ipratropium, the first-line therapy.[581] No information was provided regarding the patients' severity of disease.

As noted above, patient compliance also influences the effectiveness of therapy. One study found patient compliance with inhalers ranged from 40%-60%. Patients reported higher compliance than that found by weighing canisters. Both measures may overestimate compliance, according to the authors. Patient compliance decreased over the 5-year follow-up period.[568] Other studies have found similar rates of non-compliance.[581]

**Fosters true quality improvement**. One study examined COPD as one of three diseases possibly affected by access to care. Increased access to care in this randomized trial was associated with increased admission rates, possibly because of more detection of significant respiratory impairments in the community.[582] Thus, higher rates of COPD admission may in part reflect improvements in access to care, rather than deficient ambulatory care per se. However, this finding may also reflect a decline in the threshold for admission of "marginal" COPD cases in areas with greater access to care. It is possible that changes in coding practices, for example, coding patients as acute patients or omitting COPD codes, may reduce observed rates of COPD admissions. Recent investigations by the Medicare program into coding practices involving respiratory disease admissions raise the possibility that some COPD admissions reflect "upcoding." If the measure were used as an indicator, a decline in COPD admission rates may simply reflect a reverse change in coding practices.

**Prior use.** This measure was originally developed by Billings and colleagues in conjunction with the Ambulatory Care Project of the United Hospital Fund of New York,[275] and was subsequently adopted by the Institute of Medicine.[583] It has been widely used in a variety of studies of avoidable or preventable hospitalizations. At least 6 states (MA, NE, UT, VA, MI, NY) are reportedly using this set of measures "as guidance for policy and as evaluation and decision aids."[584] Note that COPD was not among the conditions identified by Weissman et al in 1992 as "avoidable hospital conditions."[276] A related indicator, admission for asthma, COPD, or pneumonia among patients with a prior diagnosis of COPD, was recently recommended as a measure of access to care for elderly Medicare beneficiaries.[585]

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw area level rate/standard deviation | 324.0, 203.8 | |
| Systematic area-level standard deviation* | 0.10% | High |
| Area variation as a percentage of total variation* | 0.05% | High |
| Signal ratio* | 93.4% | Very High |
| R-Square* | 93.5% | Very High |
| * age- and gender- adjusted | | |
| *Minimum Bias – APR-DRG risk adjustment* | | |
| Signal variance change with risk adjustment | Decreases | Fair |
| Absolute impact: | | |
| Average absolute change (in %) | 13.5% | Good |
| Relative impact: | | |
| Rank correlation | 0.933 | Good |
| Percent remaining in high decile/low decile | 68.2% / 86.4% | Good |
| Percent changing more than 2 deciles | 6.9% | Good |

**Precision.** This indicator is very precise, with a raw area level rate of 324.0 per 100,000 and a standard deviation of 203.8. The systematic area level standard deviation is high, at 0.10%. The area level variation also accounts for a high percentage of total variation, at 0.05%. This means that relative to other indicators, a higher percentage of the variation occurs at the area level, rather than the discharge level. The signal ratio is very high, at 93.4%. This means that it is very likely that the observed differences in area performance represent true differences in area performance. The very high R-square reflects the large proportion of signal that can be extracted using multivariate techniques. Such techniques do not have substantial additional impact, primarily due to the already very high ratio.

**Bias.** Signal variance decreases by over 15% with risk adjustment, indicating that some of the true variation among providers is due to differences in patient demographic characteristics. The indicator performs well on multiple measures of minimum bias. The rank correlation is good at 0.933. Risk adjustment appears to affect the high decile disproportionately to the low decile, as 68.2% of areas in the highest decile and 86.4% of the lowest decile remain after risk adjustment. The absolute magnitude of the impact is moderate, as is the relative impact.

**Construct validity.** COPD is related to most other ACSC indicators.

### *Discussion*

COPD can often be controlled in an outpatient setting negating the need for admission. As COPD is a chronic progressive disorder, some rate of hospitalization is appropriate. Guidelines have been established and are not widely adhered to. However, it is unclear whether adherence to these guidelines actually reduces admissions. Patient compliance has been shown to be relatively low for this condition and may influence admission rates. Access to care explains 27% of the variation in COPD admission rates. However, another study found that household income did not substantially affect admission rates. The evidence for the validity of this indicator is equivocal. Areas may wish to examine the precipitating events to admission using means such as chart review, to understand more clearly whether admissions are due to poor quality care or other problems. Examination of processes of care in outpatient settings may also illuminate the extent to which COPD rates are due to poor quality care.

This indicator is measured with high precision, as would be expected by the high number of COPD admissions. The signal ratio for this indicator was particularly high, suggesting that the high variance noted is likely to reflect true differences in performance.

Our empirical analysis identified moderate bias when risk adjusting for age and sex. In particular risk adjustment appears to affect the areas with the highest rates the most. Our literature review pointed out several more factors that may influence the progression of the disease and thus the admission rate for the disease. These include smoking and SES, and are likely to vary by area. Clinical factors that may in turn be related to behavioral risk factors that vary by area. Risk adjustment for observable characteristics, such as smoking rates, is recommended. The extent to which the progression of the disease (and thus the development of certain clinical characteristics disposing hospitalization) can be slowed by proper outpatient care was beyond the scope of this project. However, such information may be particularly helpful in understanding the relationship between quality, bias and ACSC conditions.

The admission rate for COPD is an avoidable hospitalization/ ambulatory care sensitive condition indicator. These indicators are not measures of hospital quality, but rather measures of outpatient and other healthcare. These measures would be of most interest to comprehensive health care delivery systems, such as some health maintenance organizations, or public health officials. ACSC indicators are correlated with each other and may be used in conjunction as an overall examination of outpatient care.

Areas may wish to identify hospitals that contribute the most to the overall area rate for this indicator. The patient populations served by these hospitals may be a starting point for interventions.

This indicator is recommended for inclusion in the HCUP II QI set. It received an empirical score of 17 out of 26. It is recommended with several caveats of use. As an ACSC indicator, this indicator may be viewed as a proxy for actual quality problems. As many factors influence COPD progression and hospitalization, additional risk adjustment for factors such as smoking rates may be desirable. Further, it is possible that providers may reduce admission rates without actually improving quality, by shifting care to an outpatient setting. Caution should be maintained for admission rates that are drastically below or above the average or recommended rates.

## INDICATOR 24: ACSC: CONGESTIVE HEART FAILURE (CHF) ADMISSION RATE

| | |
|---|---|
| Indicator | Area level admission rate for CHF. |
| Relationship to Quality | Proper outpatient treatment may reduce admissions for CHF. As such lower rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Admissions for CHF per 100,000 population. |
| Outcome of Interest | Discharges with ICD-9-CM principal diagnosis code for CHF in any diagnosis field (see Appendix 6)<br><br>Age 18-64 years.*<br><br>Exclude discharges with specified cardiac procedure codes (see Appendix 6) in any field.<br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |
| Population at Risk | Population in MSA or county, age 18-64 years.* |

* Rate can also be calculated for age 65 and older.

***Evidence from the literature***

**Face validity**. There are significant differences in physician management of patients with congestive heart failure, particularly depending on physician specialty (cardiologists vs. internists). [586, 587] In the community- hospital setting, the clinical practices of cardiologists are more compatible with published treatment guidelines than the clinical practices of other physicians. The benefits of cardiology specialty care include lower CHF readmission rates and better post-discharge quality-of-life measures, rather than lower mortality rates, fewer hospital charges, or shorter length of stay.[588] Despite clinical trial evidence demonstrating the effectiveness of ACE inhibitors, the drug remains under-prescribed by most physicians.[589] There are significant differences in physician management of patients with congestive heart failure, particularly depending on physician specialty (cardiologists vs. internists). [586, 587] Thus, it is plausible that such differences in community practices are reflected in differences in CHF admission rates. Because of the large numbers of patients with CHF and their substantial mortality, morbidity and cost of care, these differences may have a major impact on outcomes and health care costs.

**Precision**. Congestive heart failure is one of the leading diagnosis-related group (DRG) discharge diagnosis in the United States.[587] In the NIS, almost 3% of all discharges are for CHF, with roughly 200 discharges per hospital. Therefore, one can obtain relatively precise estimates of admission rates for CHF, although random variation may be important for small hospitals and rural areas. Billings' original study from New York reported up to 4.6-fold variation in CHF hospitalization rates, with a coefficient of variation of 0.646. [489]

**Minimum bias**. Important determinants of patient outcomes with CHF include certain demographic variables (e.g., patient age), clinical measures (e.g., left ventricular ejection fraction and serum creatinine), management issues (e.g., documentation of left ventricular function and documentation of etiology of CHF), and treatment strategies (e.g., ancillary drug use).[589] These factors appear to be correlated with socioeconomic status. Billings' original study from New York found that 59% of the substantial cross-variance was associated with differences in household income.[489] Only limited evidence exists on the extent to which such factors, rather than access to and use of high-quality medical care, accounts for differences across areas.
**Construct validity**. Some evidence suggests that access to care influences CHF hospitalization rates. Weissman et al. found CHF hospitalization rates to be variably associated with lack of insurance (adjusted RR=1.17 and 1.81 in MA and MD, respectively) and Medicaid (adjusted RR=2.41 and 2.53 in MA and MD, respectively).[276] Bindman et al. reported that self-reported access to care explained 50% of the variation in CHF hospitalization rates (e.g., more than for any other condition) at the zip code cluster level.[284] Millman et al.[491] reported that low-income zip codes had 6.1 times more CHF hospitalizations per capita than high-income zip codes in 11 states in 1988. Billings et al.[489] found that low-income zip codes in New York City (where at least 60% of households earned less than $15,000 in 1988, based on adjusted 1980 Census data) had 4.6 times more CHF hospitalizations per capita than high-income zip codes (where less than 17.5% of households earned less than $15,000). Household income explained 59% of the variation in CHF hospitalization rates at the zip code level.

**Fosters true quality improvement.** Physician practice style varies across areas, but does not explain variation in admission rates for chronic medical conditions after adjusting for community sociodemographic factors. Outpatient interventions such as the use of protocols for ambulatory management of low-severity patients, and improvement of access to outpatient care, would most likely decrease inpatient admissions for CHF.[93] There is little evidence that lower rates of CHF admission would lead to worse patient outcomes. However, practice guidelines or utilization review intended to raise physicians' threshold for admission may not be effective in reducing hospitalizations for chronic medical conditions.[496]

**Prior use.** This measure was originally developed by Billings and colleagues in conjunction with the Ambulatory Care Project of the United Hospital Fund of New York,[275] but a similar measure was developed contemporaneously by Weissman et al..[276] This measure was subsequently adopted by the Institute of Medicine,[583] and has been widely used in a variety of studies of avoidable or preventable hospitalizations. At least 6 states (MA, NE, UT, VA, MI, NY) are reportedly using this set of measures "as guidance for policy and as evaluation and decision aids."[584] Internationally, CHF admissions are tracked by the United Kingdom as part of the UK National Health Service High Level Performance Indicators.[590] A related measure is included in the DEMPAQ[591] measure set. A closely related indicator, nonelective admission for CHF, was recently recommended as a measure of access to care for elderly Medicare beneficiaries.[585]

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw area level rate/standard deviation | 521.0, 286.5 | |
| Systematic area-level standard deviation* | 0.14% | High |
| Area variation as a percentage of total variation* | 0.04% | High |

| | | |
|---|---|---|
| Signal ratio* | 93.0% | Very High |
| R-Square* | 93.2% | Very High |
| * age- and gender- adjusted | | |
| *Minimum Bias – Age-sex risk adjustment* | | |
| Signal variance change with risk adjustment | Decreases | Fair |
| Absolute impact: | | |
| Average absolute change (in %) | 19.6% | Good |
| Relative impact: | | |
| Rank correlation | 0.858 | Good |
| Percent remaining in high decile/low decile | 54.5% / 81.8% | Fair |
| Percent changing more than 2 deciles | 16.6% | Fair |

**Precision.** This indicator is very precise, with a raw area level rate of 521.0 per 100,000 and a standard deviation of 286.5. The systematic area level standard deviation is high, at 0.14%. The area level variation also accounts for a high percentage of total variation, at 0.04%. This means that relative to other indicators, a higher percentage of the variation occurs at the area level, rather than the discharge level. The signal ratio is very high, at 93.0%. This means that it is very likely that the observed differences in area performance represent true differences in area performance. The very high R-square reflects the large proportion of signal that can be extracted using multivariate techniques. Such techniques do not have substantial additional impact, primarily due to the already very high ratio.

**Bias.** Signal variance decreased by over 15% with risk adjustment, indicating that some of the true variation among providers reflects differences in patient demographic characteristics. The indicator performs fairly to well on multiple measures of minimum bias. The rank correlation is good at 0.858. Risk adjustment does appear to affect both the extremes of the distribution substantially, with only 54.5% of providers in the highest decile remaining after risk adjustment and 81.8% in the lowest decile. Further, 16.6% of providers move more than 2 deciles after risk adjustment. The absolute magnitude of the impact is moderate.

**Construct validity.** CHF is related to most other ACSC conditions.

***Discussion***

Congestive heart failure can be controlled in an outpatient setting for the most part, however, the disease is a chronic progressive disorder for which some hospitalizations are appropriate. Our literature review found some evidence of face validity in that cardiologists have lower admission rates than general practitioners and ACE inhibitors are under-prescribed. Admission rates for CHF have been associated with lack of insurance, and access to care may account for as much as 50% of the variation in admission rates. As the causes for admissions may range from poor quality care, lack of patient compliance, or problems accessing care, areas may wish to review CHF patient records to identify precipitating causes and potential targets for intervention.

Our empirical analysis showed that at a provider level this indicator is measured with high precision, with high systematic variation. The signal ratio is very high suggesting that any observed differences are likely to reflect true provider performance. Multivariate techniques do not have much additional impact; as such either univariate or multivariate smoothing is recommended.

This indicator is subject to moderate bias. Our empirical analysis (adjusting for age and sex) found that relative provider performance did change somewhat. In particular risk adjustment appears to affect the areas with the highest and lowest raw rates. Our literature review noted that

patient age, clinical measures such as heart function, and other management issues may affect admission rates. With the exception of age, few of these measures can be identified using administrative data. However, it is unclear which of these characteristics vary systematically by area, and account for bias. Areas with high rates may wish to examine the clinical characteristics of their patients to check for a more complex case mix. The extent to which the progression of the disease (and thus the development of certain clinical characteristics disposing hospitalization) can be slowed by proper outpatient care was beyond the scope of this project. However, such information may be particularly helpful in understanding the relationship between quality, bias and ACSC conditions. Some care of CHF complications occurs in emergency rooms, and would not appear in in-patient datasets. Examination of both emergency room data as well as inpatient data may give a more accurate picture of actual CHF complications rates.

Congestive heart failure admission rate is an avoidable hospitalization/ ambulatory care sensitive condition indicator. These indicators are not measures of hospital quality, but rather measures of outpatient and other healthcare. These measures would be of most interest to comprehensive health care delivery systems, such as some health maintenance organizations, or public health officials.

Areas may wish to identify hospitals that contribute the most to the overall area rate for this indicator. The patient populations served by these hospitals may be a starting point for interventions.

ACSC conditions typically vary with socioeconomic status. Examination of the SES status of an area's population, using estimates such as patient zip code or insurance status, may explain some of the area variation. However, SES is complexly related to poor access to care, so an area should not assume that none of the variance associated with SES is associated with poor access to care.

This indicator is recommended for inclusion in the HCUP II QI set. It received an empirical score of 14 out of 26. It is recommended with several caveats of use. As an ACSC indicator, this indicator may be viewed as a proxy for actual quality problems. Further, it is possible that providers may reduce admission rates without actually improving quality, by shifting care to an outpatient setting. Caution should be maintained for admission rates that are drastically below or above the average or recommended rates.

## INDICATOR 25: ACSC: DIABETES – SHORT TERM COMPLICATIONS ADMISSION RATE

| | |
|---|---|
| Indicator | Area level admission rate for short term complications of diabetes. |
| Relationship to Quality | Proper outpatient treatment and adherence to care may reduce the incidence of diabetic short term complication. As such lower rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Admissions for diabetic short term complications per 100,000 population. |
| Outcome of Interest | Number of discharges with ICD-9-CM principal diagnosis code for short-term complications (uncontrolled diabetes, ketoacidosis, hyperosmolarity, coma) per 100,000 population. (see Appendix 6)<br><br>Age 18-64 years.*<br><br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |
| Population at Risk | Population in MSA or county, age 18-64 years.* |

* Rate can also be calculated for age 65 and older.

***Evidence from the literature***

**Face validity.** Diabetic ketoacidosis, hyperosmolarity (HHNS), and coma are life-threatening complications of diabetes mellitus, particularly type 1 or insulin dependent diabetes mellitus (IDDM). Diabetic emergencies arise when there is an excess of glucose or insulin. The balance of insulin and glucose is kept by proper administration of insulin, and may involve other activities such as home blood-glucose monitoring. It has been noted in an adolescent and young adult population that better adherence to treatment (actual insulin intake vs. prescribed intake) is associated with fewer admissions for ketoacidosis and other complications.[592] Education programs for patients with diabetes have mixed results on reducing admissions for diabetic emergencies, though some have been shown to be effective.[593] It is important to note that intensive treatment (continuous insulin infusion pump, or multiple insulin injections daily) has been associated with more admissions for hypoglycemia.[594] Such intensive treatment has not been shown to have impact on admissions hyperglycemic events, but does reduce the incidence of long-term complications. Both hypoglycemic and hyperglycemic events are included in this indicator.

**Minimum bias.** Previously this indicator was defined with a hospital level denominator. Since some hospitals may be referral centers for diabetes, or may treat more difficult patients, some hospitals may have artificially high rates, though a quality problem is not present. This indicator has been redefined as an area-level measure. Some areas may have higher rates of

diabetes, due to ethnic or age composition. It would be expected that these areas would have higher admission rates for diabetic emergencies. Other factors, such as illness,[595-597] may also predispose patients to be admitted for diabetic emergencies. However, it is unlikely that any one area would experience significantly higher rates of these factors.

Admissions for diabetic emergencies can occur in both patients with existing and treated diabetes, as well as patients with previously unknown diabetes. One New Zealand study of 196 patients admitted for DKA found that 20% of admissions were new onset diabetes.[598] Two separate US studies of a US Urban African-American population found that 25% and 17% patients admitted for DKA were reportedly new onset diabetes.[595, 596]

Older age is associated with higher rates of underlying illness, more severe DKA, and better pre-hospitalization glycaemic control. This indicates that older patients may have fewer compliance issues and more complex cases.[597]

**Construct validity.** Precipitating events leading to admission may include physiologic causes, as discussed above, or the cessation of treatment due to access to care or non-compliance issues. Evidence that such causes are or are not due to access to care contributes to the construct validity of this indicator. However, such evidence has not been strongly shown. Some studies outside the US, and a few inside the US have examined the precipitating events of admission for diabetic emergencies. These studies often rely on self-report, which may be a biased measurement in and of itself. Of patients with previously known and treated diabetes, over 60% had made an error in insulin administration or had omitted insulin. Few of these patients also had underlying illness. Further, 25% of the original patients were readmitted within the 18-month study period. This study has no indication whether or not these errors were due to non-compliance, poor education, or access to care problems.[598] A Scottish study of young adult patients found that 42% of DKA admissions were due to lack of adherence to insulin treatment.[597]

In a potentially underserved population of Urban African-Americans, 2/3 of admissions were due to cessation of insulin therapy. Half of the patients stopping insulin treatment reported financial or other difficulties in obtaining insulin, while 21% reported inadequate understanding in adjusting dosages with food intake, and 14% were unsure about insulin management on sick days. Fourteen percent were clearly non-compliant. Most patients reported having been educated in diabetes care.[595] In a related study at a later date, 49% of patients with DKA, and 42% of patients with HHNS stopped or inadequately administered insulin prior the diabetic emergency.[596]

Access to care in relation to admissions has been explicitly studied and reported. Weissman[276] found that uninsured patients had a higher risk of admission for DKA and coma than privately insured patients (adjusted O.R. 2.18 – 2.77). Bindman[284] reported that an area's self-rated access to care report explained 46% of the variance in admissions for diabetes, though the analysis was not restricted to diabetic emergencies.

Several studies, including Billings[275] and Pappas,[281] showed that residents of low-income communities have a higher risk of "ambulatory care sensitive" admissions, including short-term diabetic complications, than residents of high-income communities. Of course, this is only indirect evidence of validity, because low income and high income communities may differ for many reasons other than access to care. In addition, these studies aggregated ambulatory-care sensitive admission rates across multiple conditions, so they do not clearly support the validity of component measures, such as admission rates for short-term diabetic complications. Two studies of ACSC indicators reported validation work for diabetes independent of measure sets.

Millman et al.[491] reported that low-income zip codes had 4.1 times more diabetes hospitalizations per capita than high-income zip codes in 11 states in 1988. Billings et al.[489] found that low-income zip codes in New York City (where at least 60% of households earned less than $15,000 in 1988, based on adjusted 1980 Census data) had 6.3 times more diabetes hospitalizations per capita than high-income zip codes (where less than 17.5% of households earned less than $15,000). Household income explained 52% of the variation in short term diabetes complication hospitalization rates at the zip code level.

**Fosters true quality improvement.** We found no evidence regarding the gaming of this indicator. Since diabetic emergencies are potentially life-threatening, it is unlikely that hospitals would fail to admit patients requiring hospitalization. Since this indicator is an area-level indicator, diversion to a nearby hospital is a non-issue.

**Prior use.** Admission for diabetic emergencies was included in both Billings[489] and Weissman's[276] sets of avoidable hospitalization measures. The indicator was also identified as a promising measure of quality and access by the DEMPAQ (Developing and Evaluating Performance Measures for Ambulatory Care Quality)[591] project, supported by the US Health Care Financing Administration, and the ACE (Access to Care for the Elderly) project, supported by the Physician Payment Review Commission. However, the denominator for these latter measures is limited to patients known to have diabetes, based on inpatient or outpatient claims diagnoses. This indicator, defined as a provider-level indicator, is currently an HCUP I indicator.

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw area level rate/standard deviation | 36.0. 24.6 | |
| Systematic area-level standard deviation* | 0.01% | Moderate |
| Area variation as a percentage of total variation* | 0.002% | Moderate |
| Signal ratio* | 51.7% | Moderate |
| R-Square* | 54.3% | Moderate |
| * age- and gender- adjusted | | |
| *Minimum Bias –age-sex risk adjustment* | | |
| Signal variance change with risk adjustment | No change | Good |
| Absolute impact: | | |
| Average absolute change (in %) | 0.6% | Very Good |
| Relative impact: | | |
| Rank correlation | 0.995 | Very Good |
| Percent remaining in high decile/low decile | 100% / 100% | Very Good |
| Percent changing more than 2 deciles | 0.5% | Very Good |

**Precision.** This indicator is moderately precise, with a raw area level rate of 36 per 100,000 population and a standard deviation of 24.6. The systematic area level standard deviation is moderate, at 0.01%. The area level variation also accounts for a moderate amount of total variation, at 0.002%. This means that relative to other indicators, a lower percentage of the variation occurs at the area level, rather than the discharge level. The signal ratio is also moderate, at 51.7%. This means that it is likely that some of the observed differences in area performance do not represent true differences in area performance. Multivariate techniques do not appear to improve the amount of extractable signal, as is reflected by the moderate R-square.

**Bias.** Signal variance does not change with risk adjustment. The indicator performs very well on the multiple measures of minimum bias. The rank correlation is very good at 0.995, and

risk adjustment does not appear to change the composition of providers in the highest and lowest decile. The absolute magnitude of the impact is also minimal.

**Construct validity.** Diabetes short term complications rate is related to most other ACSC indicators.

***Discussion***

The diabetic emergencies of DKA, coma and hypoglycemia arise from the imbalance of glucose and insulin. While diabetic emergencies typically arise from deviations in proper care, many emergencies occur when patients misadminister insulin or fail to follow a proper diet. Some of these instances may be attributed to lack of education or access to care problems, in addition to other reasons for non-compliance. Thus areas with high rates of diabetic emergencies may want to examine education practices, access to care and other potential causes of non-compliance when interpreting this indicator. Further information regarding precipitating events to admission may be gathered through chart review.

This indicator performed satisfactorily in the empirical analysis, and is measured with moderate precision. The systematic area level variation is moderate; this variation accounts for a moderate percentage of the total variation, relative to other indicators. The moderate signal ratio suggests that some of the observed differences between areas are not likely to reflect true differences in performance.

Risk adjustment with age and sex does not impact the relative or absolute performance of areas. Nonetheless, it is recommended that this indicator be risk adjusted, with age and sex. However, some areas may have higher rates of diabetes, due to different racial compositions and systematic differences in other risk factors. These areas may have a higher area rates, without actually having a higher proportion of individuals with diabetes developing short term complications. Risk adjustment for any of these observable characteristics, such as race, is recommended.

This indicator is somewhat unique in its definition. Factors noted in the literature review may aid users of this indicator in the interpretation of results. First, the combination of emergencies contained within this indicator, while all representing short-term complications, does represent both hyperglycemic and hypoglycemic events. Intensive therapy, used to prevent long-term complications of diabetes, actually increases hypoglycemic events, while neither increasing nor decreasing hyperglycemic events. Areas with high rates of intensive therapy may have a high rate of short-term complications arising primarily from hypoglycemic events. While intensive therapy is appropriate, high rates of hypoglycemic events with intensive treatment is not without some concern. However, areas may consider examining the rates of hyperglycemic versus hypoglycemic events, when interpreting this indicator.

Short-term diabetes admission rate is an avoidable hospitalization/ ambulatory care sensitive condition indicator. These indicators are not measures of hospital quality, but rather measures of outpatient and other healthcare, and as such are defined on an area level. These measures would be of most interest to comprehensive health care delivery systems, such as some health maintenance organizations, or public health officials. ACSC indicators are correlated with each other and may be used in conjunction as an overall examination of outpatient care.

Areas may wish to identify hospitals that contribute the most to the overall area rate for this indicator. The patient populations served by these hospitals may be a starting point for interventions.

This indicator is recommended for inclusion in the HCUP II QI set. It received an empirical score of 14 out of 26. It is recommended with two potential caveats of use. First, as an ACSC indicator, this indicator may be viewed as a proxy for actual quality problems. Second, rates of diabetes may vary systematically by area, creating bias for this indicator.

Section 3.E. Indicator 25. Diabetes Short Term Comp Admission Rate.

## INDICATOR 26: ACSC: UNCONTROLLED DIABETES ADMISSION RATE

| | |
|---|---|
| Indicator | Area level admission rate for uncontrolled diabetes. |
| Relationship to Quality | Proper outpatient treatment and adherence to care may reduce the incidence of uncontrolled diabetes. As such lower rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Admissions for uncontrolled diabetes per 100,000 population. |
| Outcome of Interest | Number of discharges with ICD-9-CM principal diagnosis code for uncontrolled diabetes, without mention of a short-term or long-term complication, per 100,000 population. (see Appendix 6)<br><br>Age 18-64 years.*<br><br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |
| Population at Risk | Population in MSA or county, age 18-64 years.* |

* Rate can also be calculated for age 65 and older.

### *Evidence from the literature*

Healthy People 2010[5] has established a goal to reduce the hospitalization rate for uncontrolled diabetes in persons 18-64 years of age to 5.4 per 10,000 population. The current reported rate is 7.2 per 10,000 population. This measure corresponds closely with the measure of short-term diabetes developed by Billings et al.[489] and evaluated and recommended in this report. The key exception is the ICD-9 codes 25002 and 25003, uncontrolled diabetes. This indicator (uncontrolled diabetes) includes only these two codes.

This indicator (uncontrolled diabetes), is intended not as a stand alone indicator, but for use with the short-term diabetes indicator. Combining the two indicators will result in the Healthy People 2010 measure. As such, this indicator was not subjected to additional literature review, as for the most part, the literature review for short-term diabetes applies.

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw area level rate/standard deviation | 34.7. 28.1 | |
| Systematic area-level standard deviation* | 0.01% | Moderate |
| Area variation as a percentage of total variation* | 0.007% | Moderate |
| Signal ratio* | 72.6% | High |

| | | |
|---|---|---|
| R-Square* | 74.2% | High |
| * age- and gender- adjusted | | |
| *Minimum Bias –age-sex risk adjustment* | | |
| Signal variance change with risk adjustment | No change | Good |
| Absolute impact: | | |
| Average absolute change (in %) | 7.9% | Good |
| Relative impact: | | |
| Rank correlation | 0.976 | Very Good |
| Percent remaining in high decile/low decile | 77.3% / 90.9% | Good |
| Percent changing more than 2 deciles | 2.3% | Very Good |

**Precision.** This indicator is moderately precise, with a raw area level rate of 34.7 per 100,000 population and a standard deviation of 28.1. The systematic area level standard deviation is moderate, at 0.01%. The area level variation also accounts for a moderate amount of total variation, at 0.007%. This means that relative to other indicators, a lower percentage of the variation occurs at the area level, rather than the discharge level. The signal ratio is high, at 72.6%. This means that it is likely that the observed differences in area performance represent true differences in area performance, though some represents noise. Multivariate techniques do not appear to improve the amount of extractable signal, as reflected by the high R-square.

**Bias.** Signal variance does not change with risk adjustment. The indicator performs well on the multiple measures of minimum bias. The rank correlation is very good at 0.976. Risk adjustment changes the composition of providers in the highest and lowest decile moderately. The absolute magnitude of the impact is also moderate.

**Construct validity.** Diabetes short term complications rate is related to most other ACSC indicators.

***Discussion***

This indicator should be used in conjunction with indicator 25. See discussion for Indicator 25: Short Term Diabetes Complications.

This indicator is recommended for inclusion in the HCUP II QI set. It received an empirical score of 14 out of 26. It is recommended with several potential caveats of use. First, as an ACSC indicator, this indicator may be viewed as a proxy for actual quality problems. Second, rates of diabetes may vary systematically by area, creating bias for this indicator. Finally, areas may reduce admission rates without improving quality, by shifting care to an outpatient setting.

## INDICATOR 27: ACSC: DIABETES – LONG TERM COMPLICATIONS ADMISSION RATE

| | |
|---|---|
| Indicator | Area level admission rate for long term complications of diabetes. |
| Relationship to Quality | Proper outpatient treatment and adherence to care may reduce the incidence of diabetic long term complications. As such lower rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Admissions for diabetic long term complications per 100,000 population. |
| Outcome of Interest | Number of discharges with ICD-9-CM principal diagnosis code for long-term complications of diabetes (renal, eye, neurological, circulatory, or complications not otherwise specified) per 100,000 population. (see Appendix 6)<br><br>Age 18-64 years.*<br><br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |
| Population at Risk | Population in MSA or county, age 18-64 years.* |

* Rate can also be calculated for age 65 and older.

***Evidence from the literature***

**Face validity.** Over 10.5 million people in the U.S. have been diagnosed with diabetes mellitus, with over 90% of those having type 2 diabetes (NIDDM).[599] Long term complications occur in the majority of diabetic patients to some degree, and include retinopathy (mild to proliferative), neuropathy, nephropathy, and microvascular disorders.[599]

Several observational studies have linked improved glycemic control to substantially lower risks of developing complications (retinopathy, neuropathy and nephropathy) in both Type 1 and Type 2 diabetes.[600] One study found that reducing glycosylated hemoglobin levels (a measure of glycemic control) has been estimated to produce over a two percentage point decrease in lifetime risk of blindness,[601] though the reduction is less for elderly patients. Another study of Type 2 diabetes reported a decrease of 60-100% in the incidence of retinopathy and macular edema, a 24%-50% decrease in gross protienuria, and a 16%-100% decrease in lower-extremity amputation, depending on the age of onset and insulin treatment.[602] Other studies have confirmed the benefit of tighter glycemic control on complications of NIDDM.[603, 604] It has been recommended that near-normal glycemic control be maintained in both NIDDM and IDDM.[605]

One mechanism of improving glycemic control is intensive therapy, which generally includes the use of a continuous insulin pump or multiple injections of insulin daily. The largest randomized control trial of intensive therapy for Type 1 or IDDM diabetes was The Diabetes

Control and Complications Trial. Intensive therapy slowed the progression and decreased the development of retinopathy by 54% and 47% respectively. Also reduced were the incidences of microalbuminuria and clinical neuropathy (39% and 69%).[606] Fours years post trial, the differences between the intensive therapy group and the control group narrowed, but remained significant.[607]

Given that appropriate adherence to therapy, and consistent monitoring of glycemic control, help to prevent complications, high-quality outpatient care should lower long-term complication rates. However, adherence to guidelines aimed at reducing complications (including eye and foot examinations, and diabetic education) has been described as modest[608, 609] with only 1/3 of patients receiving all essential services.[610]

**Precision.** Diabetes affects a large number of people, as do diabetic complications. Hospitalizations for amputations and other diabetic complications are not rare,[611] suggesting that reasonably precise estimates can be obtained. However, few studies have documented hospitalization rates for diabetic complications and the extent to which they vary across areas.

**Minimum bias.** It is possible that some sociodemographic characteristics of the population may lead to higher rates of long-term diabetic complications. Rates of diabetes are higher in Black, Hispanic and especially Native American populations than in other ethnic groups. Hyperglycemia appears to be particularly frequent among Hispanic and Native American individuals.[599] The duration of diabetes is positively associated with the development of complications. Since new-onset diabetes occurs more often in an elderly population, areas with older populations may have shorter durations of diabetes, and fewer long term complications. Though few studies have examined the validity of hospital diagnoses related to apparent long-term diabetic complications, one study found that administrative databases (including outpatient and lab data) had "disappointing" PPVs for long term complications (PPVs ranged from less than 50% to 88% for a three year period).[612] Whether such population differences and biases lead to substantial differences in diabetic complication rates across geographic areas is unclear.

**Construct validity.** As noted above, substantial evidence exists that compliance with treatment guidelines to prevent long-term complications of diabetes is low, and that long-term diabetic complications are common. However, the importance of problems in the quality of outpatient diabetes care in explaining variations in diabetic complication rates is less well understood. Compliance of physicians and patients is essential to achieve good outcomes, and it seems likely that problems with both access to and quality of care as well as patient compliance may contribute to the occurrence of complications.

**Fosters true quality improvement.** Little evidence exists on the impact of this quality improvement measure on the delivery of outpatient care for diabetes. Because the optimal hospitalization rate for this condition has not been defined, providers may decrease their rates by failing to hospitalize patients who would truly benefit from inpatient care. Although this concern cannot be dismissed, there is no published evidence of worse health outcomes in association with reduced hospitalization rates for long-term complications of diabetes. Such an effect seems implausible, given that only the most serious complications of diabetes are treated on an inpatient basis.

**Prior use.** This indicator, defined as a hospital-level indicator, is a current HCUP I indicator.

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw area level rate/standard deviation | 80.8, 58.1 | |
| Systematic area-level standard deviation* | 0.03% | Moderate |
| Area variation as a percentage of total variation* | 0.009% | Moderate |
| Signal ratio* | 75.6% | High |
| R-Square* | 76.6% | High |
| * age- and gender- adjusted | | |
| *Minimum Bias – age-sex risk adjustment* | | |
| Signal variance change with risk adjustment | No change | Good |
| Absolute impact: | | |
| Average absolute change (in %) | 12.4% | Good |
| Relative impact: | | |
| Rank correlation | 0.926 | Good |
| Percent remaining in high decile/low decile | 74.7% / 90.9% | Good |
| Percent changing more than 2 deciles | 8.3% | Good |

**Precision.** This indicator is moderately precise, with a raw area level rate of 80.8 per 100,000 population and a standard deviation of 58.1. The systematic area level standard deviation is moderate, at 0.03%. The area level variation also accounts for a moderate amount of total variation, at 0.009%. This means that relative to other indicators, a lower percentage of the variation occurs at the area level, rather than the discharge level. The signal ratio is high, at 75.6%. This means that it is likely that the observed differences in area performance represent true differences in area performance, though some is due to random noise. The high R-square reflects the high amount of signal that can be extracted using multivariate methods, though this amount is less than for other indicators. Multivariate techniques do not appear to improve the amount of extractable signal, as is reflected by the moderate R-square.

**Bias.** The signal variance does not change with age-sex risk adjustment. The indicator performs well on the multiple measures of minimum bias. The rank correlation is good at 0.926. Risk adjustment does not appear to impact the lowest or highest decile disproportionately. The absolute magnitude of the impact is moderate, and 8.3% of areas change more than two relative deciles with risk adjustment.

**Construct validity.** Long term diabetes is related to the other ACSC conditions.

### *Discussion*

Long term diabetes complications are thought to arise from sustained long-term poor control of diabetes. Intensive treatment programs have been shown to decrease the incidence of long-term complications in both type 1 and type 2 diabetes. However, it is unclear whether poor glycemic control arises from poor quality medical care, non-compliance of patients, lack of education, or access to care problems. Areas with high rates may wish to examine these factors when interpreting this indicator. Further information regarding precipitating events to admission may be gathered through chart review.

This indicator is measured with moderate precision, as shown in our empirical analysis. Given the high signal ratio, it is likely that the variation observed, reflects true differences in area performance, though some is due to random noise. Multivariate techniques do not have substantial additional impact. Therefore, either univariate or multivariate smoothing is recommended for this indicator.

Our analysis of minimum bias showed that risk adjusting by age and sex had a moderate effect of the performance of areas. The absolute impact was moderate, as well. Sociodemographic characteristics of the population, such as race, may bias the indicator, since there are higher rates of diabetes and poor glycemic control among Native Americans and Hispanic Americans. The importance of these factors as they relate to admission rates is unknown. Risk adjustment for observable characteristics, such as racial composition of the population, is recommended for this indicator. Outpatient clinics may also care for long-term complications of diabetes. Thus, examining both inpatient and outpatient data may give a more accurate picture of this indicator.

Diabetes long term complications admission rate is an avoidable hospitalization/ ambulatory care sensitive condition indicator. These indicators are not measures of hospital quality, but rather measures of outpatient and other healthcare, and as such are defined on an area level. These measures would be of most interest to comprehensive health care delivery systems, such as some health maintenance organizations, or public health officials. ACSC indicators are correlated with each other and may be used in conjunction as an overall examination of outpatient care.

Areas may wish to identify hospitals that contribute the most to the overall area rate for this indicator. The patient populations served by these hospitals may be a starting point for interventions.

This indicator is recommended for inclusion in the HCUP II QI set. It received an empirical score of 11 out of 26. It is recommended with several potential caveats of use. First, as an ACSC indicator, this indicator may be viewed as a proxy for actual quality problems. Second, rates of diabetes may vary systematically by area, creating bias for this indicator. Third, providers could reduce admission rates without improving quality of care by shifting care to an outpatient setting. Caution should be maintained for rates that are drastically below or above average rates.

## INDICATOR 28: ACSC: HYPERTENSION ADMISSION RATE

| | |
|---|---|
| Indicator | Area level admission rate for hypertension. |
| Relationship to Quality | Proper outpatient treatment may reduce admissions for hypertension. As such lower rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Admissions for hypertension per 100,000 population. |
| Outcome of Interest | Discharges with ICD-9 principal code for hypertension per 100,000 population. (see Appendix 6)<br><br>Age 18-64 years.*<br><br>Exclude discharges with specified cardiac procedure codes (see Appendix 6) in any field.<br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |
| Population at Risk | Population in MSA or county, age 18-64 years.* |

* Rate can also be calculated for age 65 and older.

***Evidence from the literature***

**Face validity.** Hypertension is a chronic condition that is often controllable in an outpatient setting with appropriate use of drug therapy. We found little literature on hypertension admission as an ambulatory care sensitive condition indicator. Hypertension was originally included in both John Billings'[489] set of indicators developed for the United Hospital Fund of New York, and in Weissman's[276] set of indicators. These sets were developed by physician panels. Evidence on sets of ambulatory care sensitive condition indicators are summarized at the beginning of this section, and should be referred to for this indicator.

**Precision.** Although hypertension is a common condition, hospitalizations for complications of hypertension are relatively uncommon. One study noted that hypertension accounted for only 0.5% of total admissions for ACSCs.[282]

**Minimum bias.** We found very little evidence on potential biases in this indicator. It is possible that the age structure of the population may affect admission rates for this condition. Weissman et al. reported a reduction of 100% in relative risk for Medicaid patients when adjusting for age and sex.[276] Though it seems plausible that differences in socioeconomic status and comorbid conditions such as obesity would affect population rates of hypertension and its complications, we found no evidence on the effects of comorbidities or other risk factors that may vary systematically by area on admission rates for hypertension complications in the area.

**Construct validity.** Two studies of ACSC conditions reported the results for hypertension independently. Bindman et al. found that an area's self rated access to care explained 22% of admissions for hypertension.[284] Weissman et al. found that uninsured patients

had a relative risk of admission for hypertension of 2.38 in Massachusetts after adjustment for age and sex, while Maryland had a corresponding relative risk of 1.93. Medicaid patients also had somewhat elevated risks (adj. RR = 1.56, 1.74).[276] Millman et al.[491] reported that low-income zip codes had 7.6 times more hypertension hospitalizations per capita than high-income zip codes in the same 11 states in 1988.

**Fosters true quality improvement.** Little evidence exists on the impact of this quality improvement measure on the delivery of outpatient care for hypertension. Because the optimal hospitalization rate for this condition has not been defined, providers may decrease their rates by failing to hospitalize patients who would truly benefit from inpatient care. Although this concern cannot be dismissed, there is no published evidence of worse health outcomes in association with reduced hospitalization rates for hypertension. Such an effect seems implausible, given that only the most serious episodes of accelerated or malignant hypertension are treated on an inpatient basis.

**Prior use.** This measure was originally developed by Billings and colleagues in conjunction with the Ambulatory Care Project of the United Hospital Fund of New York,[275] and was subsequently adopted by the Institute of Medicine.[583] It has been widely used in a variety of studies of avoidable or preventable hospitalizations. At least 6 states (MA, NE, UT, VA, MI, NY) are reportedly using this set of measures "as guidance for policy and as evaluation and decision aids."[584] This indicator was also included in Weissman's set of avoidable hospitalizations.[276]

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw area level rate/standard deviation | 37.1, 32.2 | |
| Systematic area-level standard deviation* | 0.01% | Moderate |
| Area variation as a percentage of total variation* | 0.006% | Moderate |
| Signal ratio* | 69.9% | Moderate |
| R-Square* | 71.2% | High |
| * age- and gender- adjusted | | |
| *Minimum Bias – age-sex risk adjustment* | | |
| Signal variance change with risk adjustment | No change | Good |
| Absolute impact: | | |
| Average absolute change (in %) | 9.1% | Very Good |
| Relative impact: | | |
| Rank correlation | 0.963 | Very Good |
| Percent remaining in high decile/low decile | 72.7% / 100% | Good/ V.G. |
| Percent changing more than 2 deciles | 3.2% | Very Good |

**Precision.** This indicator is moderately precise, with a raw area level rate of 37.1 per 100,000 population and a substantial standard deviation of 32.2. The systematic area level standard deviation is moderate, at 0.01%. The area level variation accounts for only a moderate percentage of total variation, at 0.006%. This means that relative to other indicators, a lower percentage of the total variation occurs at the area level, rather than the discharge level. The signal ratio is moderate, at 69.9%. This means that it is likely that some of the observed differences in area performance do not represent true differences in area performance. The high R-square denotes the high amount of signal that can be extracted using multivariate methods,

though this is less than for other indicators. Multivariate methods do not appear to have substantial additional impact.

**Bias.** Signal variance does not change with risk adjustment. The indicator performs well on multiple measures of minimum bias. The rank correlation is very good at 0.963. Risk adjustment appears to affect the highest decile disproportionately to the lowest decile, as 72.7% of areas in the highest decile and 100% of the lowest decile remain after risk adjustment. Few providers change more than 2 deciles in relative performance with risk adjustment. The absolute magnitude of the impact is minimal.

**Construct validity.** Hypertension is related to the other ACSC conditions.

***Discussion***

Hypertension is a common outpatient disorder, that can be effectively treated on an outpatient basis. Little evidence exists regarding the validity of this indicator. One study did relate admission rates to access to care problems.

Admission for hypertension is uncommon, suggesting that the indicator may be subject to some precision problems. However, our empirical results showed that this indicator is measured with adequate precision for use as a quality indicator. The high signal ratio (after multivariate smoothing) suggests that observed variation is likely to reflect true differences in quality of care, though some is also likely to reflect random noise. Multivariate techniques do not have substantial additional impact on the amount of extractable signal. As a result, either multivariate or univariate smoothing is recommended.

This indicator is subject to some minimal bias. Risk adjustment appears to impact providers with the highest rates the most, meaning that without risk adjustment, some providers may be misidentified as outliers. However, this bias is less substantial than it is for other indicators. Age and sex may be particularly important factors, and should be risk adjusted for. It is unknown how other clinical factors would impact this measure.

Hypertension is an avoidable hospitalization/ ambulatory care sensitive condition indicator. These indicators are not measures of hospital quality, but rather measures of outpatient and other healthcare. These measures would be of most interest to comprehensive health care delivery systems, such as some health maintenance organizations, or public health officials. ACSC indicators are correlated with each other and may be used in conjunction as an overall examination of outpatient care.

Areas may wish to identify hospitals that contribute the most to the overall area rate for this indicator. The patient populations served by these hospitals may be a starting point for interventions.

Overall, this indicator is recommended for inclusion in the HCUP II QI set. It received an empirical rating of 14 out of 26. This indicator is recommended with several caveats of use. As an ACSC indicator, this indicator may be viewed as a proxy for actual quality problems. Further, it is possible that providers may reduce admission rates without actually improving quality, by shifting care to an outpatient setting. Caution should be maintained for admission rates that are drastically below or above the average or recommended rates.

## INDICATOR 29: ACSC: LOWER EXTREMITY AMPUTATION RATE

| | |
|---|---|
| Indicator | Area level admission rate for lower extremity amputation. |
| Relationship to Quality | Proper and continued treatment and glucose control may reduce the incidence of lower extremity amputation. As such lower rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Admissions for lower extremity amputation in diabetics per 100,000 population. |
| Outcome of Interest | Discharges with ICD-9 procedure code for LE amputation in any field and diagnosis code for diabetes in any field per 100,000 population. (see Appendix 6)<br><br>Age 18-64 years.*<br><br>Exclude discharges with trauma (see Appendix 6).<br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |
| Population at Risk | Population in MSA or county, age 18-64 years.* |

* Rate can also be calculated for age 65 and older.

***Evidence from the literature***

**Face validity.** Lower-extremity amputation is a common complication of diabetes, affecting up to 15% of all diabetics in their lifetimes.[613] In the United States, diabetes is the leading cause of nontraumatic amputations (approximately 57,000 per year).[614] While the full etiology of factors leading to amputation are unknown, it is believed that a combination of factors contribute to the high rate of amputation in the diabetic population. Neuropathy and the subsequent loss of sensation may lead to minor trauma to the feet. These lesions, including foot ulcers, may fail to heal due to poor circulation and other factors. Resulting infections may lead to gangrene.[613, 615] Each of these singular causes can be prevented to a certain extent, leading to the prevention of lower extremity amputation. Possible interventions include foot clinics, wearing proper foot ware, and proper care of feet and foot ulcers.[615] Therefore, the American Diabetes Association (ADA) recommends that diabetics be educated in proper foot care, and that diabetics undergo a foot exam at least once a year. In addition, the Diabetes Control and Complications Trial (DCCT)[606] found that some of these factors, including neuropathy and microvascular disease, are preventable by maintaining blood glucose levels near normal using intensive insulin therapy. The DCCT has led to further recommendations that blood-glucose levels should be closely monitored in all diabetics (2-4 measurements per year), and glycosylated hemoglobin levels should be maintained below 8%.

There is substantial evidence that the ADA recommendations are not closely followed. One study of a California HMO noted that over half of patients did not have documented glycosylated hemoglobin levels (a measure of blood glucose control). In addition almost 40% of those patients with documented levels had at least one high level noted (over 10%), denoting overall poor glycemic control. Almost all patients had no documented foot exams, though these may have occurred without documentation.[616] In the median state in 1998-99, only 71% of Medicare beneficiaries had a glycosylated hemoglobin test within the previous year, while only 69% and 57% had an eye exam and a lipid profile, respectively, within the previous 2 years.[617] Several older studies, involving both Medicare[618, 619] and Medicaid[620] recipients, showed similar deficiencies in processes of care that may help prevent amputations. In addition, the Medical Outcomes Study and the National Health and Nutrition Examination Survey[618] both showed that over 50% of diabetics are less than optimally controlled, and as many as 18.8% (in staff-model HMOs) to 32.4% (in fee-for-service care) may be at extremely high risk of complications, with a glycosylated hemoglobin of 12% or more.[621]

**Precision.** Within the diabetic community, the incidence of lower extremity amputation has been reported as 375 per 100,000 person years for NIDDM, and 388 per 100,000 person years for IDDM. The twenty-five year cumulative risk for lower extremity amputation was 11%.[621]

Although we located no studies discussing small-area variation of lower extremity amputation, worldwide numbers greatly vary from 2.8 per 100,000 population (Madrid, Spain) to 43.9 in the Navajo population in the United States. The other US study site, Montgomery, AL had an age adjusted incidence of 19.2 per 100,000.[622]

**Minimum bias.** Several sociodemographic variables are associated with the risk of lower-extremity amputation in diabetics, including age, duration of diabetes, and sex.[613, 623] Males have been found repeatedly to have higher risks of amputation (2.8-6.5 fold higher rates).[613] Age and sex distribution may vary systematically by area.

Race appears to be an important factor that may influence area rates. A study of hospital discharges in 1991 in California noted that African Americans had just under twice the rate of amputation as Whites (95.25 vs. 55.98 per 10,000 persons with diabetes, RR=1.72). Hispanics had a rate similar to that among Whites (44.43 per 10,000 persons with diabetes).[624] While minorities may have up to twice the amputation rates as whites,[613] it is unknown whether this association is due to differences in access to quality care, compliance, or biological risk factors. Another study of risk factors leading to amputation noted that in an insured HMO population, African-Americans did not have greater risk of amputation,[623] suggesting that the observed differences by race are due to access to care, and should not be adjusted for. However, rates of diabetes are uniformly higher in Black, Hispanic and especially Native American populations than in other ethnic groups. Hyperglycemia appears to be particularly frequent among Hispanic and Native American individuals.[599] Therefore, adjusting or stratifying by race may be advisable when this indicator is defined using all adults, rather than all adult diabetics, as the denominator population.

Two controlled studies have identified clinical risk factors for amputation among diabetics in an HMO[623] and in a VA medical center,[625] both settings with relatively good access to care. Selby and Zhang found that the level of glucose control, duration of diabetes, and baseline systolic blood pressure were major clinical predictors of amputation. Other diabetic complications, such as microvascular complications and history of stroke, were also predictive.[623] Reiber et al. controlled for sociodemographic factors, and found that sensory

perception, circulation indices, and nutritional factors were associated with amputation risk. Patients with the most severe disease had a four-fold risk of amputation, compared to patients with less severe disease.[625] These studies suggest that some areas may have higher amputation rates than others, partially because of unmodifiable patient characteristics (e.g., duration of diabetes). However, many of the clinical factors associated with amputation are potentially modifiable **in the long term,** if excellent control of hyperglycemia and hypertension are maintained. Therefore, the magnitude of potential bias due to confounding depends somewhat on whether one takes a short-term or long-term perspective.

**Construct validity.** Several studies of intervention programs have noted a decrease in amputation risk. A 1989 prospective randomized study of a 1 hour foot care education program for high risk patients (patients with foot ulcers or previous amputation), noted a 3-fold greater risk of amputation 2 years after intervention in control patients. The education program informed patients of proper foot care and was supplemental to normal teaching regarding diet, exercise, weight and medication.[626] A more recent study noted a 1 year post-intervention decrease of 79% in amputations in a low-income African American population. The intervention varied by risk, with low risk patients receiving foot care education, and assistance in finding proper fitting footwear. High risk patients were provided with custom-molded orthoses and prescription footwear. Additional foot care was provided for patients with foot injuries.[627]

One study examining the literature noted that provider and patient education may lead to a 72% decrease in amputations, multidisciplinary clinic care, a 47% decrease, and insurance coverage for therapeutic shoes a 53.5% decrease. They calculated the potential economic benefits for the first year to be over $1.1 million, $750,000, and $850,000 for educational interventions, multidisciplinary clinics and insurance coverage for footwear respectively. However, most of the benefit would occur in individuals 70 years or older.[628]

One observational study of the risk factors of lower-extremity amputation found that patients who receive no outpatient diabetes education have a three-fold higher risk of amputation than those receiving care.[625] Although there is no clear evidence that areas with higher amputation rates provide worse care to diabetic patients, the evidence certainly suggests that high-quality care can substantially decrease amputation rates among diabetics.

**Fosters true quality improvement.** We located no evidence regarding the ability of this indicator to foster true improvement. It is unlikely, given the severity of conditions requiring lower-extremity amputations, that patients requiring amputation would be denied care.

**Prior use.** This indicator is not widely used. Healthy People 2010[5] has set a goal of reducing the number of lower extremity amputations from 4.1 per 1,000 persons with diabetes (in 1997) to 1.8 per 1,000 persons with diabetes. In addition this indicator is included in the DEMPAQ measure set for outpatient care.

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw area level rate/standard deviation | 30.5, 42.7 | |
| Systematic area-level standard deviation* | 0.04% | Moderate |
| Area variation as a percentage of total variation* | 0.001% | Moderate |
| Signal ratio* | 68.5% | Moderate |
| R-Square* | 70.2% | High |
| * age- and gender- adjusted | | |
| *Minimum Bias – age-sex risk adjustment* | | |

| | | |
|---|---|---|
| Signal variance change with risk adjustment | No change | Good |
| Absolute impact: | | |
| Average absolute change (in %) | 13.1% | Good |
| Relative impact: | | |
| Rank correlation | 0.919 | Good |
| Percent remaining in high decile/low decile | 59.1% / 90.9% | Good |
| Percent changing more than 2 deciles | 7.4% | Good |

**Precision.** This indicator is moderately precise, with a raw area level rate of 30.5 per 100,000 population and a substantial standard deviation of 42.7. The systematic area level standard deviation is moderate, at 0.04%. The area level variation accounts for only a moderate percentage of total variation, at 0.001%. This means that relative to other indicators, a lower percentage of the total variation occurs at the area level, rather than the discharge level. The signal ratio is moderate, at 68.5%. This means that it is likely that some of the observed differences in area performance do not represent true differences in area performance, though some also reflects random noise. The high R-square denotes the high amount of signal that can be extracted using multivariate methods, though this is less than for other indicators. Multivariate methods do not appear to have substantial additional impact.

**Bias.** Signal variance does not change with risk adjustment. The indicator performs well on multiple measures of minimum bias. The rank correlation is good at 0.919. Risk adjustment appears to affect the highest and lowest decile somewhat, as 59.1% of areas in the highest decile and 90.9% of the lowest decile remain after risk adjustment. The absolute magnitude of the impact is moderate.

**Construct validity.** Lower extremity amputation is slightly related to the other ACSC conditions, though it appears to be somewhat independent, as it loads on factor 2 more highly.

***Discussion***

Diabetes is a major risk factor for lower extremity risk factor. Infection, neuropathy and microvascular disease, are among the precipitating factors leading to lower-extremity amputation. Proper long term glucose control, diabetes education and foot care are just some of the interventions that have been implicated to reduce the incidence of these factors. Some observational studies have shown that high quality education and care can reduce lower extremity amputation, though no studies have reported that low quality care is associated with increased lower extremity amputation rates.

This indicator is measured with moderate precision, as shown in our empirical analysis. Given the high signal ratio (using multivariate smoothing techniques), it is likely that the variation observed, reflects true differences in area performance, though some is due to random noise. Multivariate techniques do not have substantial additional impact. Therefore, either univariate or multivariate smoothing is recommended for this indicator.

Studies have shown that lower extremity amputation varies with age and sex. Our analysis of minimum bias confirmed risk adjusting by age and sex had a moderate effect on the relative performance of areas. The absolute impact was moderate, as well. Sociodemographic characteristics of the population, such as race, may bias the indicator, since there are higher rates of diabetes and poor glycemic control among Native Americans and Hispanic Americans. However, poor quality care may also vary systematically with racial composition of the population. Therefore, it is important when adjusting for race to interpret the results with caution. Clinical risk factors such as progression of disease also affect LE amputation risk. However, the decision to include these in a risk adjustment model primarily depends on the decision to take a

long-term or short-term perspective, as progression of disease my be prevented through high quality care. Risk adjustment for observable characteristics, such as racial composition of the population, is recommended for this indicator.

The admission rate for lower extremity amputation in diabetics is an avoidable hospitalization/ ambulatory care sensitive condition indicator. These indicators are not measures of hospital quality, but rather measures of outpatient and other healthcare, and as such are defined on an area level. These measures would be of most interest to comprehensive health care delivery systems, such as some health maintenance organizations, or public health officials. ACSC indicators are correlated with each other and may be used in conjunction as an overall examination of outpatient care.

Areas may wish to identify hospitals that contribute the most to the overall area rate for this indicator. The patient populations served by these hospitals may be a starting point for interventions.

The Healthy People 2010 goal to reduce lower extremity amputation is defined with a denominator of only diabetics. The proposed indicator has a denominator of total population, as data on diabetes rates in a population is not as readily available, as census data. Nonetheless, areas with data on overall diabetes rates in the MSA or county may wish to consider this indicator in context of these rates.

Overall, this indicator is recommended for inclusion in the HCUP II QI set, though it is recommended that it be used in conjunction with other ACSC indicators.. It received an empirical rating of 10 out of 26, and smoothing is recommended. This indicator is recommended with two caveats of use. As an ACSC indicator, this indicator may be viewed as a proxy for actual quality problems. Further, this indicator has unclear construct validity, as this indicator has not been validated except as part of a set of indicators.

## INDICATOR 30: ACSC: LOW BIRTH WEIGHT RATE

| | |
|---|---|
| Indicator | Area level low birthweight rate. |
| Relationship to Quality | Proper preventative care may reduce the incidence of low birthweight. As such lower rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Number of low birth weight infants per 100 births. |
| Outcome of Interest | Number of births with ICD-9 diagnosis code for birthweight less than 2500 grams per 100 births within area. (see Appendix 6)<br><br>Exclude transfer from other institution. |
| Population at Risk | All births (discharges in MDC 15 – newborns and neonates) in MSA or county. |

***Evidence from the literature***

**Face validity.** Infants may be low birth weight due to inadequate interuterine growth or premature birth. Once premature labor has commenced, it is often quite difficult to stave the progression for any significant amount of time. However, risk factors for low birth weight may be addressed with adequate prenatal care and education. Risk factors include many sociodemographic and behavioral characteristics, such as low income and tobacco use during pregnancy. Prenatal education and care programs have been established to help reduce low birth weight and other complications in high risk populations. Addressable risks include: maternal undernutrition, genital tract infections, excessive physical exertion or stress, psychological stress and adverse health habits such as nicotine, alcohol or illicit drug exposure or poor prenatal care.[629] Nonetheless, evidence of the effectiveness of these programs has been equivocal. It is unclear whether increasing prenatal care or education actually reduces low birth weights. Healthy People 2010 has set a goal to reduce the percentage of low birth weight infants to .9%.[5]

**Precision.** Although low birth-weight births account for only a small fraction of total births, the large number of births suggest that this indicator should be precisely measurable for most areas.

**Minimum bias.** Of the risk factors for low birth weight, very few are related directly to patient care. It is unclear how many of the risk factors are in actuality indirect measures to problems accessing care. Socioeconomic measures such as parental education and income have been shown to be negatively associated with rates of low birth weight infants.[630, 631] Demographic factors such as age and race also appear important, and may be correlated with socioeconomic factors. Very young mothers (under 17 years) and older mothers (over 35 years) are at a higher risk of having low-birth weight infants.[630, 631] Other factors such as tobacco use, primiparity, complications of pregnancy or labor and delivery, and marital status have also been cited as risk factors for low birth weight.[630]

Black race has been repeatedly shown to be a risk factor for low birth weight. Many studies have attempted to control for the many confounding risk factors discussed above. One

study of all California singleton births in 1992 found that after risk adjustment having a black mother remained a significant risk factor (Adj. OR = 1.6).[630] In an attempt to delineate whether maternal birthplace affected low birth weight among black and other minority groups. David et al.[632] examined the relative risks of black mothers in Illinois born in the US, West Africa, or white mothers born in the US. They found that African born women had a 50% increase in relative risk as compared to white mothers, and US born black mothers had a 100% increase, when matching cases with respect to age, marital status, education and spouses education, prenatal care, parity and previous prenatal loss. Another study that examined California births in 1992 found no difference between foreign born and US born black women when adjusting for maternal and infant characteristics.[633] As this study adjusted for more characteristics than the David et al. study, the difference found in the David et al. study may be due to such bias. This study did find a difference between Latina US born and foreign born women, with US born women having higher rates of low-birth weight infants, even after adjustment for maternal and infant characteristics.

Low births weight also varies systematically by metropolitan versus non-metropolitan areas. An analysis of 11 million births in all 50 states from 1985-1987 found that mothers residing in non-metropolitan counties more likely to have low birthweight infants than those residing in metropolitan counties, before risk adjustment. However, after adjusting for maternal race, age less than 18, age over 35, nulliparity, parity greater than 4, single marital status, completion of high school and to a limited extent late prenatal care (this was considered an outcome as well), there were no longer statistical differences between metropolitan and non-metropolitan areas. This suggests that this indicator at an area-level could be potentially biased.[634]

The interrelationship between all these risk factors and prenatal care and complications are very complex and have not been studied in a large sample. One study of women in Baltimore noted that factors of potential social stress, such as crime rate or unemployment rate, may interact with other risk factors, creating a complicated web of risk factors.[631] Indeed, the picture of which factors lead to low birth weight is unclear because of these complex relationships.

Little evidence exists on the extent to which each of these factors contributes to differences in the rate of low birthweight births across geographic areas.

**Construct validity.** A number of studies have addressed the impact of prenatal care, or level of prenatal care among low birth weight babies. One study examined birth records in California to establish the relative risk of having a low birth weight baby as a function of use of prenatal care. Those with inadequate care (as calculated by Kotelchuck's Adequacy of Prenatal Care Use Index) had a 3.68 adjusted odds ratio of having a low birth weight infant, adjusting for maternal and infant characteristics as well as insurance status. It is important to note that those with more than adequate care had an adjusted odds ration of 6.78, with the same adjustment. These are likely to be high-risk patients that were followed closely and suggests that some of relevant risk adjustment factors were not included in this model.[630]

One randomized control trial examined the effect of prenatal care on reducing low birth weight rates in low risk women (women with past high-risk obstetrical complications, current high risk conditions, or significant comorbidities). Reducing prenatal care from 14 visits to 9 visits did not affect the rate of low birth weight infants. However, there is no indication of the socioeconomic profile of the participants and their non-clinical risk factors.[635]

Finally, one observational study, difference in the use of prenatal care accounted for less than 15% of the differences between low birth weight in black and white mothers enrolled in

Kaiser-Permanente. However, increasing level of prenatal care was associated with lower rates of low birth weight, particularly in the black patient population.[636]

One review of studies evaluating programs to reduce low birth weight notes that the studies have often been poorly designed, and this lack of rigor may account for some of the equivocal findings. The authors argue that prevention programs aimed at one specific risk factor in a population shown to have high rates of that factor do reduce low-birth rates, while comprehensive programs in potentially high risk populations do not reduce low-birth weight. They argue that this is potential evidence that prevention programs are simply misapplied and have poor designs.[629] Thus, while specific studies have demonstrated an impact of particular interventions, especially in high-risk populations, evidence on the impact of better prenatal care on low birthweight rates for area populations is less well developed.

**Fosters true quality improvement.** It seems unlikely that use of this indicator could lead to apparent reductions in the rate of low birthweight births that did not represent true reductions.

**Prior use.** Low birth weight has been used as a quality indicator on a limited basis, though interest in preventing low birth weight has been demonstrated through the literature and implementations of prevention programs. Low birth weight is a indicator in the HEDIS measure set for insurance groups, and is used by United Health Care and the University Hospital Consortium. The previous version of the HCUP I indicator set included both low birth weight and very low birth weight indicators. Healthy People 2010 has set a goal to reduce the percentage of low birth weight infants to .9%.[5]

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw area level rate/standard deviation | 3.9%, 2.3% | |
| Systematic area-level standard deviation* | 1.18% | Very High |
| Area variation as a percentage of total variation* | 0.27% | Very High |
| Signal ratio* | 67.1% | Moderate |
| R-Square* | 81.2% | High |
| * gender- adjusted only | | |
| *Minimum Bias – APR-DRG risk adjustment* | Not applicable | |

**Precision.** This indicator is precise, with a raw area level rate of 3.9% and a standard deviation of 2.3%. The systematic area level standard deviation is very high, at 1.18%. The area level variation accounts for a very high percentage of total variation, at 0.27%. This means that relative to other indicators, a higher percentage of the variation occurs at the area level, rather than the discharge level. The signal ratio is moderate, at 67.1%. This means that it is likely that some of the observed differences in area performance do not represent true differences in area performance. However, the high R-square reflects the high amount of signal that can be extracted using multivariate techniques, though this is still lower than for other indicators.

**Bias.** No analyses were conducted, since all newborns are the same age, and linkage to maternal records is not available.

**Construct validity.** Low birth weight inversely related to the other ASCS indicators. It is positively related to perforated appendix rate.

***Discussion***

Low birth weight has been implicated as an indicator of access to prenatal care. Healthy People 2010 has set a goal to reduce the percentage of low birth weight infants to .9%. However, this indicator has unclear face validity, as noted in our literature review. While mothers who give birth to low-birth weight infants generally receive less prenatal care than others and inadequate prenatal care persists as a risk factor for low birth weight when adjusting for potential confounds, comprehensive care programs in high risk women have failed to reduce low birth weight rates. It is unclear what impact the health care system has on low-birth weight. Nevertheless, potentially, and in some studies, specific counseling aimed at reducing a specific risk factor in a population identified to have that risk factor may have some impact on reducing low birth weight rates. One method of targeting populations would be to identify those hospitals contributing the most to the overall area rate. The populations served by those hospitals may be a starting place for interventions. Examination of processes of prenatal care may also help illuminate potential problem areas.

While the face validity of this indicator remains unclear, this indicator did perform well in our tests of precision. The area level standard deviation is very high. Using multivariate techniques increases the amount of signal that can be extracted for this indicator, and thus such techniques are highly recommended.

We were unable to adequately risk adjust using the data available. Adequate risk adjustment may require linkage to birth records, which record many of the sociodemographic and behavioral risk factors noted in the literature review (race, age, drug use, stress). Areas with high rates may wish to examine the prevalence of these risk factors in the population. However, some “risk factors,” while not being a direct indication of quality of care, may suggest areas of potential interventions that may reduce low-birth weight rates if implemented properly. Other “risk factors” may require some form of risk adjustment. Where risk adjustment is not possible, considering results in light of measures of SES (as determined by insurance status or patient zip code), or other factors that may provide some guidance as to “case mix” in the area. However, the relationship between potentially preventable risk factors and SES is complex and as such providers should not assume that all variance associated with SES is due to factors that cannot be influenced to reduce low birth weight. Birth records in some states are a rich source of information that could help to identify causes of low birthweight and help delineate potential areas of intervention.

Low birth weight is an avoidable hospitalization/ ambulatory care sensitive condition indicator. These indicators are not measures of hospital quality, but rather measures of outpatient and other healthcare. These measures would be of most interest to comprehensive health care delivery systems, such as some health maintenance organizations, or public health officials. ACSC indicators are correlated with each other and may be used in conjunction as an overall examination of outpatient care.

Overall, this indicator is recommended for inclusion in the HCUP II QI set, though it is recommended that it be used in conjunction with other ACSC indicators. It received an empirical rating of 11 out of 16 (bias could not be evaluated for this indicator). This indicator is recommended with several caveats of use. As an ACSC indicator, this indicator may be viewed as a proxy for actual quality problems. Further, this indicator could have substantial bias that would require additional risk adjustment from birth records or clinical data. Finally, this indicator has unclear construct validity, as this indicator has not been validated except as part of a set of indicators.

## INDICATOR 31: ACSC: PEDIATRIC ASTHMA ADMISSION RATE

| | |
|---|---|
| Indicator | Area level admission rate for pediatric asthma. |
| Relationship to Quality | Proper outpatient treatment may reduce admissions for asthma is the pediatric population. As such lower rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Admissions for pediatric asthma per 100,000 population. |
| Outcome of Interest | Number of discharges with ICD-9-CM principal diagnosis code for asthma per 100,000 population (see Appendix 6).<br><br>Age less than 18 years.<br><br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |
| Population at Risk | Population in MSA or county, age less than 18 years. |

***Evidence from the literature***

**Face validity**. Asthma is the most common chronic disease in childhood and is one of the most frequent admitting diagnoses in children's hospitals.[637, 638] In the United States, asthma affects an estimated 4.8 million children and adolescents, and in 1993 it was the cause of 198,000 admissions and 342 deaths in persons aged 24 years and younger.[637] There are effective ambulatory treatments for asthma as well as guidelines published by the National Heart, Lung, and Blood Institute, which are endorsed by the American Academy of Pediatrics.[520] These guidelines emphasize the importance of patients' access to care, appropriate diagnosis of asthma, establishment of a physician-patient relationship, timely management of asthma symptoms with appropriate medications, appropriate prophylactic and maintenance therapy, and adequate follow-up care. Healthy People 2010 has set a goal to reduce the admission rate for asthma to 25 per 10,000 population for children under 5 years, and 7.7 per 10,000 population for people aged 5-65 years.[5]

**Precision**. Because asthma is one of the most common reasons for pediatric hospitalization, with an average rate of 28.0 per 10,000 children less than 15 years of age,[637] relatively precise estimates of asthma admission across areas or hospitals can be obtained. Admission rates for asthma tend to be higher during peak times of viral respiratory infections (Winter) and allergy seasons (Spring and Fall), so care must be taken to ensure a consistent time period for measurement. There is wide variation across areas in admission rates for asthma, so random variation from year to year may be important for less populated areas

**Minimum bias**. Some admissions with asthma are unavoidable and appropriate. For example, some children have especially severe disease due to genetic factors, associated cystic fibrosis or bronchopulmonary dysplasia, or increased exposure to environmental triggers.[275, 560,

[637, 639-641] Indoor allergens such as cockroaches and dust mites may be more common in lower-income areas, and are probably associated with increased frequency and severity of asthma symptoms [523] Tobacco smoke is the most important indoor irritant and is a major precipitant of asthma symptoms in both children and adults.[520, 525-527, 529] Exposure to maternal tobacco smoke is a risk factor for the development of asthma in infancy[642] and childhood,[643-649] although not for persistence of childhood asthma into adulthood.[650]

Outdoor air pollution, especially respirable particulates, may also play a role.[520, 528-532] In addition, ozone and $SO_2$ have been associated with increased emergency department visits and hospitalizations rates.[531-538] Increasing air pollution has been specifically correlated with higher admission rates in London (Ozone, $NO_2$, $SO_2$, and black smoke),[539] and Seattle (ambient air pollution)[540].

Race represents one of the most complex potentially biasing factors for this indicator. Black patients have consistently been shown to have higher asthma admission rates,[544-546] even when stratifying for income and age.[547] One study examining differences in asthma health care utilization noted that African Americans made fewer asthma-related primary care and specialist visits than Caucasian patients (47.6% vs. 70.2% and 27% vs. 38.8%). There were no differences in hospitalization rates by race, but African-American patients had lower household incomes and made more emergency department visits (proxy for either access to care or asthma severity).[548] Similarly, Hispanics have been shown to have higher admission rates than non-Hispanic whites (or areas with higher percentages of Hispanics have been shown to have higher admission rates), although none of these studies controls for SES.[545, 549, 550] To the extent that true differences in disease prevalence or severity are responsible for racial variation in hospitalization rates, race should be adjusted for in comparing asthma hospitalization rates across areas. On the other hand, to the extent that minority patients have less access to care or poorer quality of outpatient care, race should not be adjusted for.

**Construct validity**. Little evidence has been reported conclusively attaching poor quality of care to higher area admission rates. However, numerous studies have shown that asthma hospitalization rates are associated with socioeconomic factors, including median household income (at the area level) and lack of insurance (at the individual level). A study of asthma hospitalization rates in California in 1993 (ages 0-64) found that areas with median household incomes under $35,000 had hospitalization rates that were 1.5 times higher than areas with higher median incomes.[547] In Boston, in 1992, age and gender standardized hospitalization rates (all ages) were correlated with percentage poverty in an area (r=0.68), percentage holding a bachelor's degree (r=-0.61), and income (r=-0.51).[550] Within New York City in 1994, asthma hospitalization rates were negatively correlated with a zip code area's median household income (r=-0.67), and positively correlated with the percentage of minorities in the population (r=0.82).[549] These findings confirm an earlier study by Billings et al.,[489] who reported 6.4-fold variation in asthma hospitalization rates at the zip code level in New York City in 1988, with 70% of this variation explainable by the percentage of households with annual income below $15,000. Millman et al.[491] reported that low-income zip codes had 5.8 times more asthma hospitalizations per capita than high-income zip codes in 11 states in 1988. Using New York State data, Lin et al showed that hospitalization rates were higher in areas with higher poverty, unemployment, minority populations, and lower education levels.[545] Even in England, 45% of the variation in asthma hospitalization rates across 90 family health services authorities in 1990-95 was attributable to socioeconomic factors, plus the availability of secondary care.[551] To our knowledge, only one study has reported partial correlations;[552] it found that that in New York

City, the percentage of African-American residents was the strongest predictor, and median household income was the next strongest predictor, of asthma hospitalization rates.

The observation that asthma admission rates are higher in areas with low SES has led some researchers to hypothesize that lack of access to care, or poor quality outpatient care, may lead to higher admission rates. Bindman et al.[284] showed that asthma hospitalization rates across 41 sampled areas in California were significantly correlated (r=0.47) with self-rated access to needed medical care, according to community telephone surveys. Although analyses of the National Health and Nutrition Examination Survey found that Medicaid enrollment and Spanish language preference were associated with inadequate asthma therapy, these deficiencies in care were not directly linked to hospitalizations.[554] Studies from other settings have shown that African-American asthmatics tend to have fewer scheduled primary care visits, and more hospitalizations and emergency room visits, than White asthmatics.[555, 556] African-Americans' use of asthma medications may also be less consistent with current practice guidelines.[557]

Few studies have directly linked high-quality processes of outpatient care with lower hospitalization rates at either the area or the individual level. An in-depth study of asthma treatment practices in New Haven, Boston, and Rochester found that the community with the highest asthma hospitalization rate (Boston) also had lower use of inhaled antiinflammatory agents and oral steroids. The threshold for admission also appeared to be lower in Boston, as fewer of the admitted patients were hypoxemic, relative to the other cities.[560] One case control study from a large health maintenance organization established that not having a written asthma management plan was a strong risk factor for asthma hospitalization (after adjusting for severity of asthma), but the use of anti-inflammatory medications was not.[561] With patient and parent education, good medical therapy, and outreach programs, adverse outcomes can be reduced considerably.[561, 651]

**Fosters true quality improvement.** Because the optimal hospitalization rate for this condition has not been defined, providers may decrease their rates by failing to hospitalize patients who would truly benefit from inpatient care, or by hospitalizing marginally appropriate patients with other conditions (to inflate the denominator). Although these concerns cannot be dismissed, there is no published evidence of worse health outcomes in association with reduced hospitalization rates for asthma. Indeed, given studies showing high rates of inappropriate hospitalization and poor adherence to professional guidelines, a shift to outpatient care may be entirely appropriate.[652] This is an area that should be further studied.

**Prior use.** This measure was originally developed by Billings and colleagues in conjunction with the Ambulatory Care Project of the United Hospital Fund of New York,[275] but a similar measure was developed contemporaneously by Weissman et al.[276] It was subsequently adopted by the Institute of Medicine,[583] and has been widely used in a variety of studies of avoidable or preventable hospitalizations. At least 6 states (MA, NE, UT, VA, MI, NY) are reportedly using this set of measures "as guidance for policy and as evaluation and decision aids.[276] The measure was developed to include all ages, but has since been adapted by Gadomski[277] and McConnochie[287] as a pediatric measure. South Carolina included asthma in the set of ambulatory care sensitive measures for pediatrics used by the South Carolina Department of Health Statistics.[278] Healthy People 2010 has set a goal to reduce the admission rate for asthma to 25 per 10,000 population for children under 5 years, and 7.7 per 10,000 population for people aged 5-65 years.[5]

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw area level rate/standard deviation | 154.1, 143.9 | |
| Systematic area-level standard deviation* | 0.11% | High |
| Area variation as a percentage of total variation* | 0.005% | High |
| Signal ratio* | 85.1% | High |
| R-Square* | 85.6% | High |
| * age- and gender- adjusted | | |
| *Minimum Bias –age-sex risk adjustment* | | |
| Signal variance change with risk adjustment | No change | Good |
| Absolute impact: | | |
| Average absolute change (in %) | 5.3% | Very Good |
| Relative impact: | | |
| Rank correlation | 0.994 | Very Good |
| Percent remaining in high decile/low decile | 100% / 95.5% | Very Good |
| Percent changing more than 2 deciles | 0.0% | Very Good |

**Precision.** This indicator is precise, with a raw area level rate of 154.1 and a standard deviation of 143.9. The systematic area level standard deviation is high, at 0.11%. The area level variation also accounts for a high percentage of total variation, at 0.005%. This means that relative to other indicators, a higher percentage of the variation occurs at the area level, rather than the discharge level, though still lower than for other indicators. The signal ratio is high, at 85.1%. This means that it is likely that the observed differences in area performance reflect true differences in performance, though some reflects random noise. The high R-square reflects the high proportion of signal that can be extracted using multivariate methods, though lower than for other indicators. Multivariate techniques have little additional impact.

**Bias.** The signal variance does not change with risk adjustment. The indicator performs very well on multiple measures of minimum bias. Risk adjustment does not appear to affect the extremes of the distribution substantially. The rank correlation is very good at 0.994. No of areas move more than 2 deciles in relative performance. The absolute magnitude of the impact is minimal.

**Construct validity.** Pediatric asthma is related to most other ACSC indicators.

### *Discussion*

Pediatric asthma is a chronic disease with relatively easy treatment. This indicator is related to a Healthy People 2010 goal to reduce admissions to 25 per 10,000 population age less than 5 years, and 7.7 per 10,000 population age 5-65 years. It has been noted that adherence to the guidelines for asthma management has been associated with lower admission rates, and that 71% of variance in admissions can be explained by household income.

This indicator performed well on our tests of precision. It is measured with high precision, and its signal ratio is high, suggesting that the observed variance does reflect true differences in performance. Multivariate techniques do not appear to have substantial additional impact for this indicator, and as such, either multivariate or univariate smoothing is recommended.

This indicator does not appear to be substantially biased. Risk adjustment does not appear to effect the extremes of the distribution, suggesting that without risk adjustment, areas are not likely to be mislabeled as outliers, assuming that age and sex adjustment is adequate. Our

literature review noted that some children may be at risk for admission due to comorbidities, genetic factors, and environmental triggers. It is unclear which of these factors would vary by area, nor the impact of parental compliance, which may vary systematically by area.

Pediatric asthma is an avoidable hospitalization/ ambulatory care sensitive condition indicator. These indicators are not measures of hospital quality, but rather measures of outpatient and other healthcare. These measures would be of most interest to comprehensive health care delivery systems, such as some health maintenance organizations, or public health officials.

Overall, this indicator is recommended for inclusion in the HCUP II QI set. It received an empirical rating of 18 out of 26. This indicator is recommended with several caveats of use. As an ACSC indicator, this indicator may be viewed as a proxy for actual quality problems. Further, it is possible that providers may reduce admission rates without actually improving quality, by shifting care to an outpatient setting. Caution should be maintained for admission rates that are drastically below or above the average or recommended rates.

## INDICATOR 32: ACSC: PEDIATRIC GASTROENTERITIS ADMISSION RATE

| | |
|---|---|
| Indicator | Area level admission rate for pediatric gastroenteritis. |
| Relationship to Quality | Proper outpatient treatment may reduce admissions for gastroenteritis in the pediatric population. As such lower rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Admissions for dehydration per 100,000 population. |
| Outcome of Interest | Discharges with ICD-9-CM principal diagnosis code for gastroenteritis (see Appendix 6).<br><br>Age less than 18 years.<br><br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |
| Population at Risk | Population in MSA or county, age less than 18 years. |

***Evidence from the literature***

**Face validity**. Gastroenteritis is a common illness in childhood, resulting in nearly 200,000 hospitalizations annually (nearly 10% of all admissions of children under 5 years of age).[653] There are effective ambulatory treatments for gastroenteritis and clear guidelines published both by the Centers for Disease Control and the American Academy of Pediatrics.[653] These guidelines emphasize the importance of appropriate oral rehydration therapy for mild to moderate dehydration resulting from gastroenteritis, to avoid the need for hospitalization. Adherence to these guidelines is poor; only about 33% of children treated for gastroenteritis received glucose-electrolyte solutions recommended by the AAP. Most physicians instead recommend clear liquids, which have been shown to be ineffective in treating dehydration due to gastroenteritis,[654-656] and withholding solid food longer than the 24 hours recommended by the AAP. A physician panel agreed that timely and effective ambulatory care would reduce the risk of hospitalization for gastroenteritis.[275]

**Precision**. Because gastroenteritis is one of the most common reasons for pediatric hospitalization, with small area rates of 200-400 per 100,000 children, relatively precise estimates of gastroenteritis admission across areas or hospitals can be obtained. Gastroenteritis is known to vary seasonally, with about half of hospitalizations occurring between February and April.[657] Thus the stability of the measure over time will be influenced by seasonal fluctuation in disease prevalence, so care must be taken to ensure a consistent time period for measurement. There is wide variation across areas in admission rates for gastroenteritis (14- to 18-fold differences),[657] so random variation in a particular year may be considerable for less populated areas and smaller hospitals.

**Minimum bias**. Some admissions with gastroenteritis are unavoidable and appropriate. For example, some children with gastroenteritis also suffer from a chronic disease, or from

another infection such as gingivostomatitis or tonsillitis, that inhibits oral intake of liquids.[652] These "mandatory admissions" may be difficult, if not impossible, to identify from HCUP data. However, most (73%) children admitted with gastroenteritis appear to have no underlying problems, and most (79%) are re-hydrated within 12 hours. One study suggests that complicated gastroenteritis admissions may be more common among children of low socioeconomic status.[640] If true, this finding suggests that clinical characteristics may explain some of the observed variation in admission rates for pediatric gastroenteritis.

**Construct validity**. No published studies have specifically addressed the construct validity of this indicator. Billings' original study from New York reported 1.87-fold variation in gastroenteritis hospitalization rates, with a coefficient of variation of 0.438 and 22% of variance explained by household income.[489] Millman et al.[491] reported that low-income zip codes had 1.9 times more pediatric gastroenteritis hospitalizations per capita than high-income zip codes in the same 11 states in 1988.

**Fosters true quality improvement.** Because the optimal hospitalization rate for this condition has not been defined, providers may decrease their rates by failing to hospitalize patients who would truly benefit from inpatient care, or by hospitalizing marginally appropriate patients with other conditions (to inflate the denominator). Although these concerns cannot be dismissed, there is no published evidence of worse health outcomes in association with reduced hospitalization rates for gastroenteritis. Indeed, given studies showing high rates of inappropriate hospitalization and poor adherence to professional guidelines, a shift to outpatient care may be entirely appropriate.[652] This is an area that should be further studied.

One evaluation of an intervention to improve access reported specifically on pediatric gastroenteritis. Kaestner et al. found no narrowing of the differences in "discretionary" infant (<2 year) hospitalization rates between low, middle, and high-income zip codes, during a period of substantial Medicaid eligibility expansion (1988-1992).[503] Disaggregation of gastroenteritis hospitalizations did not alter this finding.

**Prior use.** This measure was originally developed by Billings and colleagues in conjunction with the Ambulatory Care Project of the United Hospital Fund of New York.[275] It was subsequently adopted by the Institute of Medicine,[583] and has been widely used in a variety of studies of avoidable or preventable hospitalizations. At least 6 states (MA, NE, UT, VA, MI, NY) are reportedly using this set of measures "as guidance for policy and as evaluation and decision aids."[584] The measure was developed to include all ages, but has since been adapted by Gadomski[277] and McConnochie[287] as a pediatric measure. South Carolina included gastroenteritis in the set of ambulatory care sensitive measures for pediatrics used by the South Carolina Department of Health Statistics.[278]

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw area level rate/standard deviation | 98.5, 10.1 | |
| Systematic area-level standard deviation* | 0.05% | High |
| Area variation as a percentage of total variation* | 0.03% | Moderate |
| Signal ratio* | 77.8% | High |
| R-Square* | 78.8% | High |
| * age- and gender- adjusted | | |
| *Minimum Bias – age-sex risk adjustment* | | |
| Signal variance change with risk adjustment | No change | Good |

| | | |
|---|---|---|
| Absolute impact: | | |
| Average absolute change (in %) | 5.5% | Very Good |
| Relative impact: | | |
| Rank correlation | 0.995 | Very Good |
| Percent remaining in high decile/low decile | 90.9% / 100% | Very Good |
| Percent changing more than 2 deciles | 0.00% | Very Good |

**Precision.** This indicator is precise, with a raw area level rate of 98.5 and a standard deviation of 10.1. The systematic area level standard deviation is high, at 0.05%. The area level variation accounts for a moderate percentage of total variation, at 0.03%. This means that relative to other indicators, a lower percentage of the variation occurs at the area level, rather than the discharge level. The signal ratio is high, at 77.8%. This means that it is likely that the observed differences in area performance represent true variation in performance, though some may also reflect random noise. The high R-square reflects the high proportion of signal that can be extracted using multivariate techniques, though multivariate techniques do not have substantial additional impact.

**Bias.** Signal variance does not change with risk adjustment. The rank correlation is very good at 0.995. Risk adjustment does not appear to impact extremes of the distribution substantially, as 90.9% of areas in the highest decile and 100% of the lowest decile remain after risk adjustment. No areas move more than two deciles in relative performance. The absolute impact is also minimal, with an average change in performance relative to the mean of 5.5%.

**Construct validity.** Pediatric gastroenteritis is related to most other ACSC indicators.

***Discussion***

Pediatric gastroenteritis can be treated on an outpatient basis. Guidelines for this condition have been established, yet are not widely adhered to. However, there is little compelling evidence that adherence to these guidelines reduces admission rates. In fact, many of the admissions appear to be discretionary and possibly inappropriate admissions. Our literature review noted that 22% of variance in admission rates is explained by household income. Areas may wish to examine several factors when interpreting the results of this indicator. Admissions may be precipitated by poor quality care, lack of compliance with care, or poor access to care or may be due to environmental causes. Areas may wish to examine the causes of admissions through means such as chart review. The appropriateness of admissions may also be examined, to ascertain whether the admission threshold is lower in one area than another. Examination of processes of care in outpatient settings may also illuminate the extent to which gastroenteritis rates are due to poor quality care.

This indicator is measured with precision, and the signal ratio is high. This suggests that the observed variance is likely to reflect true differences in provider performance. Multivariate techniques do not appear to improve substantially the amount of signal that can be extracted, and thus either univariate or multivariate smoothing is recommended.

We did not identify substantial bias in our empirical analyses. Our literature review identified socioeconomic status to be a large factor in admission rates. Parental compliance, and increases in discretionary admissions with low parental coping resources may also influence admission rates. Most of these factors could not be identified using administrative data, and may vary systematically by area. Areas with high rates may want to identify disease severity by looking at the degree of dehydration of patients and comorbidities to establish whether or not admissions are discretionary, appropriate or due to poor quality care.

Pediatric gastroenteritis is an avoidable hospitalization/ ambulatory care sensitive condition indicator. These indicators are not measures of hospital quality, but rather measures of outpatient and other healthcare. These measures would be of most interest to comprehensive health care delivery systems, such as some health maintenance organizations, or public health officials. ACSC indicators are correlated with each other and may be used in conjunction as an overall examination of outpatient care.

Areas may wish to identify hospitals that contribute the most to the overall area rate for this indicator. The patient populations served by these hospitals may be a starting point for interventions.

Admission for pediatric gastroenteritis, like many ACSC conditions, typically varies with socioeconomic status. Examination of the SES status of an area's population, using estimates such as patient zip code or insurance status, may explain some of the area variation. However, SES is complexly related to poor access to care, so an area should not assume that none of the variance associated with SES is associated with poor access to care.

Overall, this indicator is recommended for inclusion in the HCUP II QI set, though it is recommended that it be used in conjunction with other ACSC indicators.. It received an empirical rating of 17 out of 26. This indicator is recommended with several caveats of use. As an ACSC indicator, this indicator may be viewed as a proxy for actual quality problems. This indicator has unclear construct validity, as this indicator has not been validated except as part of a set of indicators. Further, it is possible that providers may reduce admission rates without actually improving quality, by shifting care to an outpatient setting. Caution should be maintained for admission rates that are drastically below or above the average or recommended rates.

## 3.E.5. In-Hospital Mortality Measures

**INDICATOR 33: ACUTE MYOCARDIAL INFARCTION (AMI) MORTALITY RATE**

| | |
|---|---|
| Indicator | Provider level mortality rate for AMI. |
| Relationship to Quality | Better processes of care may reduce mortality for AMI. As such lower rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Number of deaths per 100 discharges with diagnosis code for AMI. |
| Outcome of Interest | Number of deaths with diagnosis code for AMI (see Appendix 6) in any field. |
| Population at Risk | All discharges with diagnosis codes for AMI in any field. (see Appendix 6)<br><br>Age 18 years and older.<br><br>Exclude transfers to other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |

***Evidence from the literature***

**Face validity**. Acute myocardial infarction (AMI) affects 1.5 million people each year and approximately one-third die in the acute phase of the heart attack.[658] Many clinical and observational studies have been conducted showing processes of care linked to survival improvements. These research findings have resulted in detailed practice guidelines covering all phases of AMI management.[486] Starting in 1992, the Health Care Financing Administration implemented a national initiative to gather data for quality improvement. The project, "the Cooperative Cardiovascular Project (CCP)," focuses on improving treatment of AMI patients.

**Precision**. The precision of AMI mortality rate estimates may be problematic for medium and small hospitals. About 13% of AMI patients in the California Hospital Outcomes Project died during hospitalization, or within 30 days of admission.[659] Since 19.5% of AMI patients were transferred from their initial hospital to another acute care facility, the percentage of deaths in an unlinked episode of care would have been somewhat less. In a study using the 1992 MedisGroups Comparative Database with 100 hospitals, mostly in Pennsylvania and the southern United States, the in-hospital mortality rate was 13.2% for AMI patients. In elderly Medicare AMI patients, 30-day mortality rates varied from 18% in Connecticut to 23% in Alabama.[660] Although these outcome rates are high, the number of AMI patients varies widely across hospitals, based on the size and risk profile of each hospital's catchment area.

**Minimum bias.** Starting in 1990, ICD-9-CM included a fifth digit for AMI codes to distinguish treatment during the "initial episode of care" from subsequent treatment related to the same AMI (within 8 weeks of the event). In studies comparing chart and administrative data since this time, the agreement in identification of new AMI cases has been shown to be at least 93%, and as high as 98%.[659, 661] The California Hospital Outcomes Project found that "unlikely" AMI patients had significantly higher mortality than patients with definite or possible AMI.[659] However, there was no evidence of systematic bias across hospitals; high-mortality and low-mortality hospitals had similar proportions of "unlikely" AMI patients.

About 19.5% of AMI patients were transferred from the initial hospital to another acute care facility in the California Hospital Outcomes Project, and these transfer rates varied across hospitals.[659] In studies using unlinked data, hospitals transferring a large proportion of their AMI patients may have lower death rates than hospitals that do not regularly transfer patients. A related bias results from the fact that many deaths related to AMI occur after hospital discharge, but within 30-days. [662] Thus, as described below (under "Fosters True Quality Improvement"), hospitals with long mean LOS may appear to have higher mortality rates than hospitals with shorter mean LOS. Investigating hospital LOS and transfer rates in conjunction with AMI mortality may help resolve these concerns.

*Risk adjustment*. Numerous studies have established the importance of risk adjustment for AMI patients. As a result, researchers have developed a number of risk adjustment models. Normand et al developed and validated two models, one of which was based on conditions likely to be present on admission and therefore applicable to comparisons of hospital-based care.[84] The claims-based model included 25 comorbidities not related to treatment. Hypertension (18.3%), diabetes (13.8%), and pulmonary disease (11.2%) were the most frequent comorbidities in an AMI Medicare cohort of 164,427 patients. Examples of frequent comorbidities that were considered possibly related to hospital treatment, and therefore omitted from their model, included congestive heart failure (33.9%), chronic angina (27.4%), and arrhythmias (25.2%). The same team developed another model using the clinical predictors available from the Cooperative Cardiovascular Project.[84] From these and numerous other studies, the most important predictors of short-term AMI mortality have been shown to include age, previous AMI, tachycardia, pulmonary edema and other signs of congestive heart failure, hypotension and cardiogenic shock, anterior wall and Q-wave infarction, cardiac arrest, and serum creatinine or urea nitrogen. Fewer studies have addressed whether adjusting for potential complications as well as comorbidities, or adjusting only for predictors available from administrative data, leads to bias in comparisons across hospitals.

Krumholz et al compared seven models including a newly developed 7-variable clinical/demographic risk adjustment model for 30-day mortality in AMI patients.[123] The models based on clinical data demonstrated better discrimination and calibration than two models[660] based on ICD-9-CM codes (area under the receiver operating curve 0.74-0.78 versus 0.70-0.71, respectively). In addition, the clinical models classified hospital performance somewhat differently than the models based on administrative data. Such differences were further explored by Iezzoni and colleagues, who used several proprietary products to estimate risk-adjusted AMI mortality, and found 40-60% disagreement in identifying the 10 best and 10 worst hospitals in a nationwide sample.[9, 204] Adding full clinical data to administrative data for risk-adjustment, Pine found that 73% of Cleveland hospitals' expected mortality rates changed by less than one standard deviation, and 100% changed by less than two.[19] In St. Louis, 95% of hospitals' expected mortality rates changed by less than 0.5 standard deviations, and 100% changed by less

than one. These estimates were better than those for other major medical conditions, including pneumonia, stroke, and congestive heart failure.[205] In the California Hospital Outcomes Project, the addition of clinical risk factors to a reestimated model based on reabstracted ICD-9-CM codes had a minimal effect on the difference in risk-adjusted mortality between low-mortality and high-mortality hospitals, although individual hospitals were affected.[659] In summary, these studies found that the method of risk-adjustment does affect which specific hospitals are identified as mortality outliers, but that the correlations within pairs of risk-adjusted or expected mortality rates are generally high (e.g., 0>0.80)[12] to 0.94[205], and higher for AMI than for other medical conditions.

When risk adjustment models include ICD-9-CM conditions that may represent consequences of poor care, then discrimination is exaggerated.[123] Romano and Chan compared an administrative data set to a re-abstraction of diagnoses present at admission, with two versions of the All Patient Refined-Diagnosis-Related Groups (APR-DRG), Risk of Mortality (ROM) and Severity of Illness (SOI).[113] The authors showed empirically that APR-DRGs predicted 30-day mortality better when all diagnoses were included than when only diagnoses present at admission were included. Hospitals' expected mortality rates based on all reabstracted ICD-9-CM codes were moderately correlated (r=0.72-0.77) with expected mortality rates based only on diagnoses present at admission. However, 2 of the 3 hospitals classified as having higher than expected mortality, 8 of the 23 hospitals classified as having neither higher nor lower than expected mortality, and 0 of the 4 hospitals classified as having lower than expected mortality, switched categories when diagnoses not present at admission were excluded from risk-adjustment.

**Construct validity**. Numerous randomized controlled trials have conclusively demonstrated that early administration of aspirin and thrombolytic agents can reduce AMI mortality.[663-667] Similarly, early revascularization by percutaneous coronary angioplasty reduces mortality in high-risk patients.[668-670] Angiotensin converting enzyme inhibitors reduce mortality among post-infarction patients with impaired left ventricular function.[671-673] Therefore, there is clear evidence at the patient level that specific processes of care improve patient outcomes. Furthermore, numerous studies based on large regional or national samples have shown substantial practice variation in AMI patients, with underutilization of clearly beneficial therapies and overutilization of harmful treatments.[95, 674, 675]

Over the last several years, substantial evidence for construct validity at the hospital level has emerged. In the first study of this type, Park et al. estimated the contribution of differences in severity of illness and quality of care to the classification of some hospitals as having unexpectedly high inpatient death rates (age and gender adjusted).[38] Not unexpectedly, severity of illness (using chart data) accounted for some of the variation. However, a quality score derived from an explicit set of process measures did not explain differences between low-mortality and high-mortality hospitals. In fact, the relationship was in the opposite direction from the authors' expectation under several analysis scenarios.

More favorable evidence came from Meehan and colleagues, who evaluated coding accuracy, severity of illness, and process-based quality of care in Connecticut hospitals.[661] Three process measures were selected by an expert panel based on medical literature and local practice patterns: 1) administration of thrombolytic therapy, 2) discharged on aspirin if no contraindication, and 3) discharged on a beta blocker if no contraindication. The hospitals with the highest risk-adjusted mortality had significantly lower utilization of beneficial therapies than the other hospitals in the sample. Although the Medicare Prospective Payment System Quality of Care study did not focus on specific therapeutic interventions, it also demonstrated

significantly higher risk-adjusted mortality rates (using risk factors derived by chart review) among hospitals with "poor" processes of care than among hospitals with "good" or "medium" processes of care (30.1% versus 22.0% and 23.9%, respectively).[676] Chen [146] showed that the hospitals designated by US News and World Report as "America's Best Hospitals" in cardiology, based on risk-adjusted mortality (using APR-DRGs) and reputation among physicians, had lower risk-adjusted mortality (using clinical predictors) among Medicare patients (15.6% versus 18.3-18.6%) and used aspirin and beta blockers more often than hospitals that were not so designated. Similarly, major teaching hospitals in the same Medicare data set had 20% lower risk-adjusted 30-day mortality than nonteaching hospitals; about half of this difference was attributable to greater use of beneficial therapies.[677] In the RAND PPS Quality of Care study in 1990, patients with higher process scale scores for AMI demonstrated significantly lower risk-adjusted 30-day AMI mortality on four out of five subscales and on an overall process scale.[676] In the California Hospital Outcomes Project, hospitals with low risk-adjusted AMI mortality were more likely to give aspirin within 6 hours of arrival in the emergency room, more likely to perform cardiac catheterization and revascularization procedures within 24 hours, and more likely to give heparin to prevent thromboembolic complications. However, there were no differences between low and high-mortality hospitals in the use or timing of thrombolytic or beta blocker therapy.[659]

These somewhat conflicting findings may relate to the general insensitivity of mortality rates to process measures. Mant and Hicks conducted a systematic review of the literature to estimate the effect sizes for therapies proven effective for AMI patients, based on clinical trials and meta-analyses.[43] The therapies assessed were beta blockade, aspirin, fibrinolysis, and angiotensin converting enzyme inhibitors. Using the best estimates of effect size and the proportion of patients eligible for treatment, the authors calculated the absolute risk reduction for low and high baseline mortality situations, with a resulting range of 5.1% to 16.4%. Given this range, they simulated the number of patients required to detect differences in care using either a "perfect system" for risk-adjusted mortality or a process-based quality of care audit. Using the same population of AMI patients, the difference in lives lost was detectable with one year of data collection on mortality or only two weeks of data collection on process of care.

The widespread recognition of the exceptionally strong evidence base supporting specific processes of care for AMI patients has led to numerous professional guidelines, guideline implementation projects,[678, 679] and regional and national quality improvement initiatives. Through its Cooperative Cardiovascular Project and Sixth Scope of Work, the Health Care Financing Administration has focused considerable attention on improving processes of care for AMI, as a way to improve mortality and other outcomes. Hospitals in the four pilot states involved in this project (AL, CT, IA, WI) significantly improved their performance on each process indicator between 1992 and 1995, and simultaneously achieved a greater reduction in 30-day mortality (19.9% to 17.6%) than hospitals in other states (19.6% to 18.2%). This finding suggests, but does not prove, that hospitals can lower their AMI mortality rates by improving adherence to evidence-based guidelines.

**Fosters true quality improvement**. In general, physicians and hospitals have little discretion in their decisions to admit AMI patients, so it seems unlikely that the use of this indicator would impede access to needed care. However, a few patients who fail to respond to, or are ineligible for, resuscitative efforts in the emergency room may not be admitted if there is pressure to reduce inpatient mortality. Although such practices might bias comparisons of risk-adjusted inpatient mortality across hospitals, they would be unlikely to compromise patient

outcomes (as resuscitative measures that fail in an emergency room would also fail in a coronary care unit). It is conceivable that patients could be discharged early to die at home or in a nursing home, although this may be unlikely due to the acute nature of the condition. Patient transfers to other hospitals will also have a greater effect on inpatient mortality rates, as noted in the OSHPD study, because hospitals vary widely in their transfer rates. Typically, 30-day overall mortality rates and 30-day inpatient mortality rates have been considered more valid than inpatient mortality rates based only on the initial hospitalization for AMI. The rank correlation between standardized AMI mortality measures based on inpatient deaths and measures based on 30-day deaths (at the hospital level) was 0.79 in a study of Medicare data.[289] This finding suggests that changes in length of stay may modestly alter the ranking of hospital performance using this measure.

**Prior use.** Inpatient AMI mortality, based on administrative data, has recently been used as a hospital quality indicator by the University Hospital Consortium,[370] the California Hospital Outcomes Project,[680] HealthGrades.com,[377] the Michigan Hospital Association (aggregated with congestive heart failure and angina),[373] and the Greater New York Hospital Association.[372] In addition, the following organizations have used this indicator with risk-adjustment based on clinical data obtained through review of medical records: the Pennsylvania Health Care Cost Containment Council[681] and Cleveland Health Quality Choice.[374] The Joint Commission for the Accreditation of HealthCare Organizations has adopted AMI mortality (from the MEDSTAT Corporation) as one of its core hospital performance measures.[443] AMI mortality is also a High-Level Performance Indicator for the United Kingdom's National Health Service.[590]

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw provider level rate/standard deviation | 24.4%, 16.1% | |
| Systematic provider -level standard deviation** | 3.4% | High |
| Provider variation as a percentage of total variation** | 0.8% | High |
| Signal ratio** | 42.8% | Moderate |
| R-Square** | 59.0% | Moderate |
| **APR-DRG, age-, gender- adjusted | | |
| *Minimum Bias – APR-DRG risk adjustment* | | |
| Signal variance change with risk adjustment | Decreases | Fair |
| Absolute impact: | | |
| Average absolute change (in %) | 32.4% | Fair |
| Relative impact: | | |
| Rank correlation | 0.747 | Fair |
| Percent remaining in high decile/low decile | 36.3% / 67.3% | Fair |
| Percent changing more than 2 deciles | 29.0% | Fair |

**Precision.** This indicator is precise, with a raw provider level mean of 24.4% and a standard deviation of 16.1%. The systematic provider level standard deviation is high, at 3.4%. The provider level variation also accounts for a high percentage of total variation, at 0.8%. This means that relative to other indicators, a higher percentage of the variation occurs at the provider level, rather than the discharge level, although more of the variation occurs at the discharge level than some indicators. The signal ratio is only moderate, at 42.8%. This means that it is likely that the some of the observed differences in provider performance do not represent true

differences in provider performance. The moderate R-square (59%) reflects the higher proportion of signal that can be extracted using multivariate techniques.

**Bias.** Signal variance decreases by over 25% with risk adjustment, indicating that some of the true variation among providers reflects differences in patient characteristics. The indicator performs fairly on the multiple measures of minimum bias. The rank correlation is fair at 0.747. The impact on the extremes is large. Only 36.3% of providers in the highest decile remain, and only 67.3% in the lowest decile remain, after risk adjustment. Similarly, the number of providers moving at least two deciles in relative rank is also high. The absolute magnitude of risk adjustment is also substantial.

**Construct validity.** AMI mortality does not load substantially on any of the three extracted factors. However, AMI mortality is correlated with several other indicators, including Bi-lateral catheterization ($r=-.16$, $p<.0001$), mortality for CHF ($r=.46$, $p<.0001$), pneumonia ($r=.46$, $p<.0001$), CABG ($r=.50$, $p<.0001$), stroke ($r=.40$, $p<.0001$), and GI hemorrhage ($r=.38$, $p<.0001$).

***Discussion***

Reductions in the mortality rate for acute myocardial infarction on both the patient level and the provider level has been related to better processes of care. Timely and effective treatments are essential for patient outcome, and include appropriate use of thrombolytic therapy and when appropriate, revascularization. The evidence surrounding the validity of AMI mortality as a quality indicator is substantial.

AMI mortality rate is measured with adequate precision, with high systematic variation, and a moderate signal ratio. This suggests that some of the observed variance may not actually reflect true differences in performance. Multivariate techniques help in extracting additional signal and are recommended. Using smoothed estimates (MSX) may help to avoid precision problems due to random noise.

Risk adjustment may be important for this indicator. Our empirical results show substantial impact of risk adjustment, particularly at the extremes. This means that without risk adjustment, some providers may be mislabeled as outliers. In addition, some of the potential risk adjustment factors, such as clinical measures, may not be available using administrative data. Methods such as chart review may help illuminate the need for more detailed risk adjustment, and potential case-mix differences between providers. Since AMI is an urgent medical condition, this indicator is not expected to be subject to selection bias.

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

Overall, this indicator is recommended for inclusion in the HCUP II QI set. It received an empirical rating of 5 out of 26, and smoothing is highly recommended. This indicator is recommended with two major caveats of use. Thirty-day mortality may be significantly different than in-hospital mortality, leading to information bias. Second, risk adjustment for clinical factors, or at minimum APR-DRGs, is recommended due to the confounding bias for this indicator.

## INDICATOR 34: CONGESTIVE HEART FAILURE (CHF) MORTALITY RATE

| | |
|---|---|
| Indicator | Provider level mortality rate for CHF. |
| Relationship to Quality | Better processes of care may reduce short-term mortality for CHF. As such lower rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Number of deaths per 100 discharges with principal diagnosis code of CHF. |
| Outcome of Interest | Number of deaths with principal diagnosis code for CHF (see Appendix 6). |
| Population at Risk | All discharges with principal diagnostic code of CHF (see Appendix 6).<br><br>Age 18 and older.<br><br>Exclude discharges with cardiac procedure codes in any field (see Appendix 6).<br>Exclude transfers to other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |

***Evidence from the literature***

**Face validity.** Admission for heart failure is common with 400,000 new cases[682] and 274,000 deaths[683] each year in the United States. Approximately 2 million persons in the U.S. have heart failure and this number will increase as the population ages.[682] Population data from the United States have not indicated a decline in mortality over the last 20 years.[684] However, a recent study from Scotland has suggested that case-fatality mortality rates have improved between 1986 and 1995.[685]

The accuracy of ICD-9 coding for heart failure has been questioned. Although the specificity of a principal diagnosis of heart failure is high (>95) the sensitivity is low.[686] Even when the principal or secondary diagnoses are used the sensitivity is only 63%, and the positive predictive value is 83.5%[686] Others have found lower positive predictive values (62.5%) but higher sensitivities (89.9%) for the combined use of principal and secondary diagnoses.[687] Face validity will be maximized by limiting analyses to patients with a principal diagnosis of heart failure.

**Precision.** Rates of short-term mortality vary from 6%[688] to 13%, [689] 14%,[690] 19.9%.[685]

In-hospital survival appears to have improved recently. The 3-day mortality rate decreased by 41% in a study of 29,500 elderly patients in Oregon from 1991 to 1995.[691] These data suggest that hospitals have improved care for heart failure patients.

There is substantial variation in hospital survival. In an analysis of 6 hospitals the survival ranged from 3.6% to 11.3%; however, no hospital had a statistically significantly higher rate.[692]

**Minimum bias.** Mortality from heart failure is greatly influenced by patient characteristics. Mortality has been reported to be higher in older patients, males and[693] and possibly whites.[694] Using administrative data, the c-statistic was 0.68 for logistic models of in-hospital survival using the Charlson comorbidity (non-specific) score and 0.78 for a heart failure specific score.[695] The variables associated with increased in-hospital mortality in the heart failure specific score were age (relative weight 1), transfer,[683] cerebrovascular disease,[683] chronic obstructive pulmonary disease,[683] hyponatremia,[683] other hydro-electrolytic disturbance,[683] metastatic disease,[683] moderate to severe renal disease,[685] ventricular arrhythmia,[687] mild liver disease,[687] malignancy,[687] hypotension and shock.[694, 695]

Another multivariable model of in-hospital mortality using clinical data (blood pressure, electrocardiogram) demonstrated a c-statistic of 0.9.[692] When six hospitals were evaluated using this clinical model, expected mortality varied from 4% to 9% with one hospital having a significantly lower predicted rate (P = .01). The observed-expected mortality ranged from -3.8% to +4.7% (all NS).

**Construct validity.** There were no studies that specifically examined the construct validity of in-hospital mortality from heart failure. We did identify several studies that provide information on the validity of this indicator. On a patient level processes of care have been shown to decrease mortality; it is unknown how implementing these processes of care would actually affect provider-level mortality rates.

Survival is known to be improved for patients with heart failure and low ejection fraction if they receive ACE inhibitors.[696] There is a wide variation in the use of ACE inhibitors during hospitalization.[697] A measure of left ventricular ejection fraction (e.g. echocardiography) is recommended to determine which patients have depressed ventricular function and would benefit from life-prolonging medical therapy. Measures of left ventricular function have also been found to vary widely within hospitals.[697] Use of echocardiography has been associated with more ACE inhibitor use and improved survival.[698] In their PPS Quality of Care study in 1990, RAND reported better process of care for CHF on all five subscales and overall. On four of five process subscales and on the overall scale, patients demonstrated significantly lower risk-adjusted CHF mortality. For patients with a poor overall process scale score (in the lowest 25%), 30-day risk-adjusted mortality was 19% while that for patients with medium process scale scores was 13% and for patients with high scale scores (in the highest 25%) 30-day risk-adjusted mortality was 11%.[676]

**Fosters true quality improvement.** Risk adjusted measures of mortality may lead to an increase in coding of comorbidities. Patients may be discharged early to a lower level of care (nursing home) so that the death does not occur in-hospital. All in-hospital mortality measures may create perverse incentives to reduce hospital mortality by discharging patients earlier, and thereby shifting deaths to skilled nursing facilities or outpatient settings. This phenomenon may also lead to biased comparisons among hospitals with different mean lengths of stay. The rank correlation between standardized mortality measures based on inpatient deaths and measures based on 30-day deaths (at the hospital level) was 0.71 and 0.78 in studies of Medicare[289] and all-payer Cleveland[699] data, respectively. This finding suggests that changes in length of stay may modestly alter the ranking of hospital performance using this measure. However, Rosenthal

et al. found no evidence that hospitals with lower in-hospital standardized mortality had higher (or lower) early post-discharge mortality.

**Prior use.** Mortality for congestive heart failure has been widely used as a quality indicator. In addition to the use in the literature listed above, HealthGrades.com,[377] the University Hospital Consortium[370] and the Greater New York Hospital Association[372] have used this measure. The Maryland Hospital Association include this measure in their Maryland QI Project Indicator set.[369] Cleveland Health Quality Choice[374] includes CHF mortality in their cardiovascular care measure. Likewise, Michigan Hospital Association[373] includes CHF in an aggregated mortality measure.

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw provider level rate/standard deviation | 7.5%, 9.5% | |
| Systematic provider -level standard deviation** | 2.1% | Moderate |
| Provider variation as a percentage of total variation** | 0.7% | Moderate |
| Signal ratio** | 53.5% | Moderate |
| R-Square** | 69.7% | Moderate |
| **APR-DRG, age-, gender- adjusted | | |
| *Minimum Bias – APR-DRG risk adjustment* | | |
| Signal variance change with risk adjustment | No change | Good |
| Absolute impact: | | |
| Average absolute change (in %) | 13.7% | Good |
| Relative impact: | | |
| Rank correlation | .794 | Good |
| Percent remaining in high decile/low decile | 39.8% / 72.9% | Fair |
| Percent changing more than 2 deciles | 23.5% | Fair |

**Precision.** This indicator is precise, with a raw provider level mean of 7.5% and a standard deviation of 9.5%. The systematic provider level standard deviation is moderate, at 2.1%. The provider level variation also accounts for a moderate percentage of total variation, at 0.7%. This means that relative to other indicators, a lower percentage of the variation occurs at the provider level, rather than the discharge level. The signal ratio is only moderate, at 53.5%. This means that it is likely that the some of the observed differences in provider performance do not represent true differences in provider performance. The moderate R-square (69.7%) reflects the higher proportion of signal that can be extracted using multivariate techniques.

**Bias.** Signal variance does not change with risk adjustment. The indicator performs fairly to well on the multiple measures of minimum bias. The rank correlation is good at 0.794. The impact on the extremes is large. Only 39.8% of providers in the highest decile remain, and only 72.9% in the lowest decile remain, after risk adjustment. Similarly, the number of providers moving at least two deciles in relative rank is also high. The absolute magnitude of risk adjustment is moderate.

**Construct validity.** Congestive heart failure mortality loads on factor 1. It is positively related to other medical mortality indicators, such as pneumonia, GI hemorrhage, and stroke, and to a lesser extent, hip fracture.

***Discussion***

Congestive heart failure is a progressive chronic disease. Short-term mortality is substantial and varies from provider to provider. Certain treatments have been shown to decrease short-term mortality on a patient level, but it is unknown whether or not such practices decrease provider-level mortality.

CHF is a relatively common admission, with a relatively high short-term mortality rate. Our empirical tests showed that this indicator is precise, with moderate systematic variation, and a moderate percentage of that variation is provider-level. The signal ratio is moderate, suggesting that the some of the differences observed do not reflect true differences in performance. Multivariate techniques improve the amount of signal that can be extracted for this indicator and as such MSX smoothed estimates are recommended to avoid problems with precision due to random noise.

While CHF mortality has not been studied extensively as an indicator, some have developed risk models for short-term death. Comorbidities and some clinical factors appear important in predicting death. Our empirical analysis confirmed that risk adjustment impacts this indicator. Relative performance is the most impacted, particularly for providers with high and low rates. Therefore, it is important to use some sort of risk adjustment for this indicator, as otherwise, some providers may be mislabeled as outliers. Another source of bias is the outpatient management of some patients. Some providers may admit only the most severely ill patients and handle other patients in an outpatient setting, while others do not do this. This results in a more severe casemix. Providers may wish to examine rates of outpatient care.

Overall, this indicator is recommended for inclusion in the HCUP II QI set. It received an empirical rating of 6 out of 26, and smoothing is highly recommended. This indicator is recommended with three major caveats of use. First, some CHF care occurs in an outpatient setting, and selection bias may be a problem for this indicator. Second, thirty-day mortality may be significantly different than in-hospital mortality, leading to information bias. Third, risk adjustment for clinical factors, or at minimum APR-DRGs, is recommended due to the confounding bias for this indicator.

## INDICATOR 35: GASTROINTESTINAL (GI) HEMORRHAGE MORTALITY RATE

| | |
|---|---|
| Indicator | Provider level mortality rate for GI hemorrhage. |
| Relationship to Quality | Better processes of care may reduce mortality for GI hemorrhage. As such lower rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Number of deaths per 100 discharges with principal diagnosis code of GI hemorrhage. |
| Outcome of Interest | Number of deaths with principal diagnosis code for GI hemorrhage (see Appendix 6). |
| Population at Risk | All discharges with gastrointestinal hemorrhage in principal diagnosis field (see Appendix 6).<br><br>Age 18 years or older.<br><br>Exclude transfers to other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates |

***Evidence from the literature***

**Face validity**. Admission for gastrointestinal hemorrhage is fairly common (100/100,000 adults). Mortality rates for hemorrhage vary greatly, and lower mortality has been associated with more use of treatments such as early endoscopy (within 24-48 hours of presentation), though the strength of this relationship has not been established, with some studies failing to find significant relationships (see construct validity section). Mortality rates in large population based databases have not changed since the 1940s, though, there have been increases in the ages and comorbidities of patients.[700]

**Precision**. Rates of mortality in gastrointestinal hemorrhage vary from 0-29%, with most studies reporting rates of 3.5%-11%.[700-704]

**Minimum bias**. Mortality from gastrointestinal hemorrhage is highly influenced by patient comorbidities and other factors complicating the bleed, as well as the nature and severity of the bleed itself, which all vary substantially across patients with the condition. One study noted that some endoscopic findings, hemodynamic characteristics, and comorbidities were highly predictive of life-threatening events.[705] The same study found at reassessment that the strongest predictors of life-threatening events were reoccurring bleeding (3-6 times risk of mortality) and unstable comorbid diseases. Mortality rates in patients with gastrointestinal hemorrhage have been shown to increase with age. In one study, the overall mortality, with the absence of comorbidity, was 4%, while patients under the age of 60 years experienced a mortality of 0.1%. Concurrent malignancy and organ failure increased the mortality rate for patients under 60 years to 0.8%.[704]

Patients who develop bleeding in-hospital have higher mortality rates than patients admitted with gastrointestinal bleeding that began outside the hospital. One study reported a difference in crude mortality rates of 33% versus 11%. [702, 704]

One study tested the effect of risk-adjustment on hospital ranking for gastrointestinal hemorrhage mortality. Risk-adjusting for age, shock, and comorbidity (characteristics that are often reported on discharge abstracts) changed 30 hospitals rankings by more than 10. Adding diagnosis, endoscopy findings, and rebleed status changed 32 hospital rankings by more than 10.[700]

**Construct validity**. We located no studies explicitly evaluating the construct validity of this indicator. On a patient level processes of care have been shown to decrease mortality. However, it is unknown how implementing these processes of care would actually affect provider-level mortality rates.

A number of medical treatments have been shown to be associated with bleeding control, though evidence on association with mortality is more limited. Endoscopy has been shown inconsistently to be associated with mortality. One meta-analysis showed a slight advantage for early endoscopy,[706] while another study found that endoscopy was not related to mortality in either the bivariate or multivariate analyses.[701]

Many of the deaths reported are not associated with bleeding per se. One study found that only one death was related to bleeding, and that patient had several severe comorbidities.[702] Thus, in many cases, the deaths in patients with a diagnosis of gastrointestinal hemorrhage are probably not actually due to the bleed itself.

**Fosters true quality improvement**. Risk-adjusted measures of mortality may lead to an increase of coding of comorbidities or upcoding of diagnoses. All in-hospital mortality measures may create perverse incentives to reduce hospital mortality by discharging patients earlier, and thereby shifting deaths to skilled nursing facilities or outpatient settings. This phenomenon may also lead to biased comparisons among hospitals with different mean lengths of stay. We found no published evidence about whether the difference between inpatient and 30-day mortality for in GI hemorrhage is substantial enough to cause concern.

**Prior use.** GI hemorrhage is currently used by the Cleveland Choice Health Quality Choice[369] and the Maryland Hospital Association (in the Maryland QI Project).[369] GI hemorrhage is also included in the Michigan Hospital Association's aggregated mortality measure.[373]

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw provider level rate/standard deviation | 4.6%, 5.7% | |
| Systematic provider -level standard deviation** | 1.1% | Moderate |
| Provider variation as a percentage of total variation** | 0.3% | Moderate |
| Signal ratio** | 20.2% | Low |
| R-Square** | 55.5% | Moderate |
| **APR-DRG, age-, gender- adjusted | | |
| *Minimum Bias – APR-DRG risk adjustment* | | |
| Signal variance change with risk adjustment | Decreases | Fair |
| Absolute impact: | | |
| Average absolute change (in %) | 10.5% | Good |
| Relative impact: | | |

| | | |
|---|---|---|
| Rank correlation | 0.803 | Good |
| Percent remaining in high decile/low decile | 48.9% / 32.2% | Fair |
| Percent changing more than 2 deciles | 35.5% | Fair |

**Precision.** This indicator is precise, with a raw provider level mean of 4.6% and a standard deviation of 5.7%. The systematic provider level standard deviation is moderate, at 1.1%. The provider level variation also accounts for a moderate percentage of total variation, at 0.3%. This means that relative to other indicators, a lower percentage of the variation occurs at the provider level, rather than the discharge level. The signal ratio is low, at 20.2%. This means that it is very likely that the some of the observed differences in provider performance do not represent true differences in provider performance. The moderate R-square (55.5%) reflects the higher proportion of signal that can be extracted using multivariate techniques.

**Bias.** Signal variance decreases by more than 25% with risk adjustment, suggesting that some of the observed variance is due to differences in patient characteristics. The indicator performs fairly to well on the multiple measures of minimum bias. The rank correlation is good at 0.803. The impact on the extremes is large. Only 48.9% of providers in the highest decile remain, and only 32.2% in the lowest decile remain, after risk adjustment. Similarly, the number of providers moving at least two deciles in relative rank is also high. The absolute magnitude of risk adjustment is moderate.

**Construct validity.** GI hemorrhage mortality loads on factor 1. It is positively related to mortality indicators, such as pneumonia, stroke, and congestive heart failure, and to a lesser extent, hip fracture.

***Discussion***

GI Hemorrhage may lead to death when uncontrolled. However, our literature review noted that the bleed itself is rarely the cause of death, calling into question the face validity of this indicator as a measure of quality of care for hemorrhage. The ability to manage severely ill patients with comorbidities may influence the mortality rate for "GI hemorrhage," though we found no evidence of this hypothesis.

GI hemorrhage mortality rate is measured with adequate precision, with moderate provider systematic variation. Though mortality due to the bleed itself is low, mortality in patients with GI hemorrhage along with other comorbidities is high enough and varies enough to expect adequately precise measurement. The signal ratio for this indicator is low, suggesting that some of the observed differences likely do not reflect true differences in performance. However, multivariate techniques do improve the amount of signal that can be extracted, and are recommended. Using smoothed estimates (MSX) may help to avoid precision problems due to random noise.

The extreme influence of comorbidities on the survival rate of patients with GI hemorrhage, as well as the influence of age and timing of onset (pre or post hospitalization) raises questions about the potential bias for this indicator. While we found no published evidence that these factors vary systematically by provider it seems likely that hospitals may vary in their treatment of geriatric or severely ill patients. Our empirical analysis confirmed this potential bias, particularly when identifying overall changes in provider performance or the providers with the lowest and highest mortality rates. Providers should risk adjust for comorbidities. In addition, providers with high rates may want to examine their case-mix for higher complexity of cases (patients over 60, more comorbidities).

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

Overall, this indicator is recommended for inclusion in the HCUP II QI set. It received an empirical rating of 5 out of 26, and smoothing is highly recommended. This indicator is recommended with two caveats of use. First, risk adjustment for clinical factors, or at minimum APR-DRGs, is recommended due to the substantial confounding bias for this indicator. Second, limited evidence supports the construct validity of this indicator.

**INDICATOR 36: HIP FRACTURE MORTALITY RATE**

| | |
|---|---|
| Indicator | Provider level mortality rate for hip fracture. |
| Relationship to Quality | Better processes of care may reduce mortality for hip fracture. As such lower rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Number of deaths per 100 discharges with principal diagnosis hip fracture. |
| Outcome of Interest | Number of deaths with principal diagnosis code for hip fracture (see Appendix 6). |
| Population at Risk | All discharges, with principal diagnosis code of hip fracture (see Appendix 6).<br><br>Age 18 or older.<br><br>Exclude transfers to other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |

***Evidence from the literature***

**Face validity**. Hip fractures are a common cause of morbidity and functional decline among elderly persons. In addition, hip fractures are associated with a significant increase in the subsequent risk of mortality, which persists for a minimum of 3 months among the oldest and most impaired individuals at baseline,[707, 708] and perhaps up to several years among younger and less impaired individuals.[709, 710] Fractures of the femoral neck or intertrochanteric region are usually caused by minimal trauma (e.g., fall on a step, on a level surface, or from a chair) in the setting of osteoporosis, a common condition characterized by demineralization and weakening of weight-bearing bones. About 89% of hip fracture patients are elderly, and they often have multiple comorbidities and pre-fracture functional impairments. As a result, they are at significant risk of such postoperative complications as pneumonia, myocardial ischemia, arrhythmias, and deep vein thrombosis. If these complications are not recognized and effectively treated, life-threatening problems such as respiratory failure, myocardial infarction, and pulmonary embolus may ensue.

**Precision**. Hip fracture is the most common type of fracture requiring hospitalization; about 382,000 discharges in 1998 listed a diagnosis of hip fracture, of which 329,000 listed it as the principal diagnosis (12.0 per 10,000 persons).[711] Based on the all-payer database in California, each hospital admitted an average of 51.6 elderly hip fracture patients requiring surgical repair per year in 1995-96.[712] The largest published study of in-hospital mortality reported a rate of 4.9% in 1979-88.[713] These data suggest that mortality rates are likely to be relatively reliable at the hospital level.

**Minimum bias**. There is relatively little potential for selection bias, because almost all patients with hip fracture are hospitalized. However, hip fracture patients may tend to be sicker at some hospitals than at others, due to variations in health status across the catchment areas of different facilities. The known predictors of in-hospital or 30-day mortality can be divided into several categories. Demographic predictors include age, male sex, and prior residence in a nursing home. Comorbidity predictors include malnutrition, venous disease, digestive diseases, cardiovascular diseases (including congestive heart failure, angina, and atrial fibrillation), neoplasms, disorientation or delirium, chronic obstructive pulmonary disease, prior hospitalization within one month, the Charlson comorbidity index (or the number of chronic medical conditions), and the ASA (American Society of Anesthesiology) physical status score. The only proven functional predictor of short-term mortality is dependency in any activity of daily living. Finally, fracture site may be a significant predictor, although probably more for long-term outcomes than for short-term outcomes. To the extent that these factors are more prevalent at some hospitals than at others, hip fracture mortality rates may be susceptible to confounding bias. In the absence of studies explicitly comparing models with and without clinical data elements, it is difficult to assess whether administrative data contain sufficient information to remove bias.

**Construct validity**. There is conflicting evidence on the construct validity of this indicator. The association between risk-adjusted mortality (using clinical data obtained by chart abstraction) and implicit and explicit process criteria was explored as part of RAND's Prospective Payment System Quality of Care study. Whereas Medicare patients with poor "process of care" had higher risk-adjusted 30-day mortality than those with good "process of care" for four medical conditions, there was no difference for hip fracture (4.6% versus 5.1%, RR=0.90). None of the process subscales (physician cognitive, nurse cognitive, technical diagnostic, technical therapeutic, monitoring) was associated with risk-adjusted 30-day mortality.[676] A more recent British study identified one East Anglican hospital with significantly lower than average risk-adjusted mortality (at 90 days); patients at this hospital "were routinely treated by a designated multidisciplinary team for fracture of the hip, with early assessment and surgery, much of which was performed by one surgeon," routine thromboembolism prophylaxis, and early mobilization.[714]

There is very little evidence supporting an association between hospital volume and mortality following hip fracture repair. (Following Halm, Lee, and Chassin[80], we did not find this evidence to be sufficiently strong to recommend total hip fracture volume as a separate volume indicator.) Using administrative data from Florida, without any risk-adjustment, Lavernia[715] found no association between surgeon volume and in-hospital mortality. They did not report the effect of hospital volume, if any. A study of Medicare data from 1979 and 1980 showed no association with 60-day mortality, after adjusting for age, sex, and region,[716] although a more recent study of 1988 and 1990 Medicare data found higher than expected in-hospital mortality at low-volume hospitals.[717] Maerki[718] and Luft[239] found "no clear pattern" in the relationship between hospital volume and hip fracture mortality, using 1972 data from the Commission on Professional and Hospital Activities (CPHA), although there was a suggested effect among low risk patients.[719, 720] Hughes and colleagues[52] used 1982 data from CPHA and did find a significant association between volume and inpatient mortality in a hospital-level regression analysis. These inconsistencies provide limited support for the construct validity of mortality as a quality indicator.

There is also substantial evidence that at least two major causes of death among hip fracture patients are partially preventable. Perez et al.[721] reviewed 581 autopsy reports on hip fracture patients who died in a single British hospital between 1953 and 1992, and reported that 80 (14%) and 55 (9%) deaths were attributable to pulmonary emboli and acute myocardial infarction, respectively. The most common cause of death after hip fracture was bronchopneumonia (46%); no medical intervention has been shown to reduce the incidence of, or mortality due to, this complication. Thromboembolic prophylaxis using either unfractionated or low molecular weight heparin reduces the incidence of radiographically documented deep vein thrombosis from 39% to 24% (OR=0.41, 95% CI=0.31-0.55), whereas physical devices reduce the incidence from 19% to 6% (OR=0.24, 95% CI=0.13-0.44).[722] Although meta-analysis suggests a reduction in fatal pulmonary emboli with heparin prophylaxis (OR=0.39, 95% CI 0.14-1.09), there are insufficient data to draw firm conclusions regarding symptomatic emboli. A recent controlled trial suggests that aspirin reduces the incidence of symptomatic deep vein thrombosis (RR=0.71, 95% CI=0.52-0.97), symptomatic pulmonary emboli (RR=0.57, 95% CI=0.40-0.82), and fatal pulmonary emboli (RR=0.42, 95% CI=0.24-0.73) after hip fracture.[723] These experimental data are supported by population-based observational data from at least two areas.[714, 724] One randomized controlled trial that included high-risk patients undergoing noncardiac surgery suggested that perioperative use of beta blockers may reduce the incidence of postoperative AMI.[725] A recent meta-analysis reported a nonstatistically significant reduction in AMI (RR=0.70, 95% CI=0.64-3.57) and a marginally significant reduction in total mortality at one month (RR=0.72, 95% CI=0.51-1.00) with regional anesthesia.[726, 727] Nutritional supplementation may be a useful strategy to prevent postoperative complications, but no effect on mortality has been demonstrated.[728] Finally, several aspects of surgical technique may be associated with higher short-term mortality, including the use of hemiarthroplasty instead of internal fixation for displaced femoral neck fractures,[729-732] the use of cement[731] or a posterior approach[732] in hemiarthroplasty for such fractures, and delayed surgical fixation.[729, 730, 733-735] However, these findings come from observational studies that are susceptible to bias. The deleterious effect of hemiarthroplasty disappears with longer follow-up.[732, 736] Nonetheless, these studies give providers with high hip fracture mortality rates some ideas for investigation or intervention.

**Fosters true quality improvement**. One possible adverse effect of in-hospital mortality measures is to encourage earlier postoperative discharge. This is a definite concern, for several reasons. First, 30-day mortality for hip fracture is substantially higher than in-hospital mortality (4.6-6.8% versus 3.3-4.9% in the largest published studies), suggesting that a relatively modest decrease in mean length of stay could significantly decrease inpatient mortality. Second, mean length of stay decreased more dramatically for hip fracture (i.e., by 5.6 days) than for any of the other four conditions studied after Medicare introduced prospective payment.[737] This decrease was associated with a disproportionate decrease in risk-adjusted in-hospital mortality (from 5.7% to 3.3%, versus a decrease in 30-day mortality from 5.3% to 4.6%) and an increase in instability at discharge (from 18.8% to 23.1%), which was correlated with higher post-discharge mortality.[738] There is conflicting evidence about whether the decrease in mean LOS has led to more prolonged nursing home stays.[737, 739-742] Another potential response would be to avoid operating on high-risk patients, although there is such consensus in favor of surgical management of hip fracture that avoidance of high-risk patients seems unlikely.

**Prior use.** In-hospital mortality following hip fracture repair has not been widely used as a quality indicator, although it is subsumed within a University Hospital Consortium indicator

(mortality for DRG 209).[370] The United Kingdom National Health Service also uses hip fracture mortality as a High Level Performance Indicator.[590]

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw provider level rate/standard deviation | 14.4%, 16.0% | |
| Systematic provider -level standard deviation** | 7.8% | High |
| Provider variation as a percentage of total variation** | 6.0% | Very High |
| Signal ratio** | 54.3% | Moderate |
| R-Square** | 65.0% | Moderate |
| **APR-DRG, age-, gender- adjusted | | |
| *Minimum Bias – APR-DRG risk adjustment* | | |
| Signal variance change with risk adjustment | No change | Good |
| Absolute impact: | | |
| Average absolute change (in %) | 22.5% | Fair |
| Relative impact: | | |
| Rank correlation | 0.880 | Good |
| Percent remaining in high decile/low decile | 55.6% / 78.8% | Good/ Fair |
| Percent changing more than 2 deciles | 15.1% | Fair |

**Precision.** This indicator is precise, with a raw provider level mean of 14.4% and a standard deviation of 16.0%. The systematic provider level standard deviation is high, at 7.8%. The provider level variation also accounts for a very high percentage of total variation, at 6.0%. This means that relative to other indicators, a higher percentage of the variation occurs at the provider level, rather than the discharge level. The signal ratio is only moderate, at 54.3%. This means that it is likely that the some of the observed differences in provider performance do not represent true differences in provider performance. The moderate R-square (65%) reflects the higher proportion of signal that can be extracted using multivariate techniques.

**Bias.** Signal variance does not change with risk adjustment. The indicator performs well to fairly on the multiple measures of minimum bias. The rank correlation is moderate at 0.880, and risk adjustment does seem to impact disproportionately the lowest decile relative to the highest decile, compared to other indicators. Only 78.8% of providers in the low decile without risk adjustment remain after risk adjustment. Similarly, 15.1% of providers change more than 2 deciles in performance.The absolute magnitude of the impact of risk adjustment is substantial.

**Construct validity.** Hip fracture mortality loads on factor 1. It is positively related to other medical mortality indicators, including pneumonia, stroke, GI hemorrhage, and CHF mortality.

### *Discussion*

Hip fracture occurs in frequently in the elderly population. Complications of fracture and treatments sometimes include embolism, pneumonia, and myocardial ischemia. These conditions and other comorbidities lead to a relatively high mortality rate, and there is some evidence that some of these complications are preventable.

Such high mortality rates suggest that this indicator will be measured with good precision. Our empirical analyses found that this indicator is very precise, with high systematic variation. However, the signal ratio is moderate, suggesting that some of the observed variance

does not reflect true differences in performance. Multivariate techniques improve the ability to extract signal for this indicator, and as such smoothing using MSX or other methods is advised.

Though all patients with hip fracture are hospitalized, some specialty centers may admit more clinically severe or frail patients. Patient age, sex, comorbidities, fracture site, and functional status are all predictors of functional impairment, though there is little evidence that these vary systematically by hospital. Our empirical analyses confirmed that this indicator has some potential bias. The impact of APR-DRG risk adjustment is moderate to substantial for both absolute and relative performance. Risk adjustment with age and sex and APR-DRGs is highly recommended. Further, detailed chart reviews may identify differences in functional status or other clinical factors not accounted for in discharge data.

Hospital discharge practices differ, with some hospitals discharging patients earlier than others, and some transferring patients to rehabilitation or sub-acute care facilities. For this reason, this indicator should be considered in conjunction with length of stay and transfer rates (though transfers are excluded in this indicator).

Overall, this indicator is recommended for inclusion in the HCUP II QI set. It received an empirical rating of 10 out of 26, and smoothing is highly recommended. This indicator is recommended with three caveats of use. First, thirty-day mortality may be somewhat different than in-hospital mortality, leading to information bias. Second, risk adjustment for clinical factors, or at minimum APR-DRGs, is recommended due to the confounding bias for this indicator. Third, there is limited evidence for the construct validity of this indicator.

## INDICATOR 37: PNEUMONIA MORTALITY RATE

| | |
|---|---|
| Indicator | Pneumonia mortality rate |
| Relationship to Quality | Inappropriate treatment for pneumonia may increase mortality |
| Benchmark | State, regional or peer group average |

Method:

| | |
|---|---|
| Quality Measure | Mortality in discharges with principal diagnosis code of pneumonia |
| Outcome of Interest | Number of deaths with principal diagnosis code per 100 discharges |
| Population at Risk | All discharges with principal diagnosis code for pneumonia (see Appendix 6).<br><br>Age 18 years and older.<br><br>Exclude transfers to other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates |

***Evidence from the literature***

**Face validity**. Pneumonia constitutes the sixth leading cause of death in the United States,[743] causing approximately one million adult hospitalizations per year and approximately 75,000 in-hospital deaths, according to the NIS database. Patient characteristics are relatively important predictors of in-hospital mortality, though the performance of specific processes of care may also lead to better patient outcomes.

**Precision**. Although pneumonia admissions are common, the high degree of patient heterogeneity suggests that pneumonia mortality indicators will be imprecise. One study[39] examined the impact of choice of statistical methods for handling random variation in hospital mortality rates for pneumonia in patients under 65 years of age. Using the prognostic model (MedisGroups™) employed by a managed care company in reporting mortality for 22 Pennsylvania hospitals, the authors reanalyzed the data adjusting for multiple comparisons and other analytic errors produced by "simplistic" statistical methods. Their analysis indicated that the "simplistic" analysis resulted in a 60% chance that at least one hospital would be incorrectly labeled a high outlier for mortality.

**Minimum bias**. Hospitals and area physicians vary in their threshold to admit patients with pneumonia, with some admitting more "low risk" patients than others.[93] Comparison of hospital death rates with population death rates suggests that selection bias due to such differences in admission practice influences observed hospital mortality rates for pneumonia.[76] Hospitals may also transfer pneumonia patients with severe chronic comorbid illnesses to nonacute facilities or home for their terminal care. Thus, population death rates from pneumonia (in particular, non-inpatient deaths) may be an important supplement to indicators based on hospital mortality.

*Risk adjustment*. Generic risk adjustment models do not perform well as disease specific models for pneumonia mortality.[744] The variables contained in the pneumonia PORT scoring system in predicting mortality[121] suggest that administrative data generally do not contain enough information for highly accurate risk adjustment. Important predictors of outcome not reliably captured in administrative databases include the microbial etiology,[745] certain radiographic patterns [745, 746] and pre-hospital functional status.[747] Comparisons of models that include additional clinical risk adjusters with administrative risk adjustment bear out these concerns.[13, 18]

**Construct validity**. While the impact of patient factors on hospital mortality rates for pneumonia must be emphasized, it is also true that processes of care contribute to pneumonia outcomes, including mortality. However, it is unknown how implementing these processes of care, which have been shown to affect patient level outcomes, would actually affect provider-level mortality rates. One mechanism is through the choice of antibiotics. In their PPS Quality of Care study in 1990, RAND reported better process of care for pneumonia on all five subscales and overall. On four of five process subscales and on the overall scale, patients demonstrated significantly lower risk-adjusted pneumonia mortality. For the technical therapeutic process subscale, mortality varied from 15% (good process scale score, i.e. in the highest 25%) to 21% (low process scale score, in the lowest 25%) mortality.[676] A recent study reported an association between choice of antibiotics and 3-day mortality for patients hospitalized with pneumonia.[748] Compared to recommended regimens involving second or third-generation cephalosporins, use of beta-lactam/beta-lactamase inhibitors plus a macrolide, or an aminoglycoside plus another agent were associated with an increased 30-day mortality. Based on their spectrum of antimicrobial activity, the choice of regimens associated with increased mortality might reflect concern on the part of treating physicians for greater severity of illness. The authors used a previously validated, clinically based prognostic model[121] to adjust for differences in patient risk of death. Risk-adjustment using this model did not suggest greater severity of illness for patients on the non-recommended regimens. Other results in the study, however, suggest that elements of patient risk remain unaccounted for by this model [749] (e.g., patients on the non-recommended regimens had a higher rate of ICU treatment within 24 hours of hospital arrival, which seems likely to reflect disease severity and not quality of care). Confounding by indication likely explains the results of another recent study of macrolide use in the treatment of community-acquired pneumonia (CAP).[750]

More basic than the choice of a particular antibiotic regimen is the timely administration of any antibiotic to the patient presenting to the hospital with community-acquired pneumonia. A retrospective study of Medicare patients with pneumonia showed that patients who received antibiotics within 8 hours and blood cultures within 24 hours of presentation had significantly lower mortality than patients for whom these processes were not performed.[751] Timely delivery of antibiotics bears a plausible connection to improved outcomes. However, the infrequency with which blood cultures alter therapy for patients with pneumonia suggests that this association reflects the performance of other correlated but hard-to-measure aspects of care. Although a causal connection in the case of rapid performance of blood cultures is unlikely, these findings do suggest that some quality problems existed at the hospitals with higher mortality rates.

Probably for similar reasons, structural measures that seem plausibly associated with better processes of care are also associated with better outcomes. For instance, teaching status correlated with higher quality scores for treatment of pneumonia in one study.[218] The quality scores, however, involved compliance with process measures only, not outcome differences. A

different study that did examine outcomes reported that teaching hospitals achieved no better risk-adjusted 30-day mortality, but also incurred higher costs.[752]

Reflecting the limitations of the evidence on the impact of therapy on pneumonia outcomes, relatively few practice guidelines have focused on reducing mortality. Rather, guidelines for pneumonia are more likely to be concerned with the implications of antibiotic choice for future drug resistance.[753, 754] These guidelines aim to improve long-term community outcomes due to lower prevalence of resistant microorganisms along with lower costs of treatment represent the goals of pneumonia.[748, 751]

**Fosters true quality improvement**. One study successfully improved timely delivery of antibiotics to patients with pneumonia,[755] but the intervention was not prompted by mortality data, nor did mortality due to pneumonia change as a result of the intervention.

All in-hospital mortality measures may create perverse incentives to reduce hospital mortality by discharging patients earlier, and thereby shifting deaths to skilled nursing facilities or outpatient settings. This phenomenon may also lead to biased comparisons among hospitals with different mean lengths of stay. Although we found no data on the sensitivity of standardized pneumonia mortality measures (at the hospital level) to the inclusion or exclusion of post-discharge deaths, inpatient mortality among Medicare patients in California was only 77% of 30-day mortality in 1985.[106] This gap has presumably expanded over the last 15 years. This finding suggests that changes in length of stay may modestly alter the ranking of hospital performance using this measure.

**Prior use.** Pneumonia mortality is currently widely discussed in the literature. In addition it is a commonly used mortality indicator. Users include: the University Hospital Consortium,[370] HealthGrades.com,[377] Greater New York Hospital Association,[372] Maryland Hospital Association (as part of the Maryland QI Project),[369] the Pennsylvania Health Care Cost Containment Council and the California Hospital Outcomes Project.[681] In addition, the Michigan Hospital Association includes pneumonia in an aggregated mortality measure.[373]

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw provider level rate/standard deviation | 13.8%, 10.2% | |
| Systematic provider -level standard deviation** | 3.7% | High |
| Provider variation as a percentage of total variation** | 1.2% | High |
| Signal ratio** | 62.9% | Moderate |
| R-Square** | 78.1% | High |
| **APR-DRG, age-, gender- adjusted | | |
| *Minimum Bias – APR-DRG risk adjustment* | | |
| Signal variance change with risk adjustment | Decreases | Moderate |
| Absolute impact: | | |
| Average absolute change (in %) | 21.9% | Fair |
| Relative impact: | | |
| Rank correlation | 0.596 | Fair |
| Percent remaining in high decile/low decile | 30.6% / 60.3% | Fair |
| Percent changing more than 2 deciles | 40.8% | Fair |

**Precision.** This indicator is precise, with a raw provider level mean of 13.8% and a standard deviation of 10.2%. The systematic provider level standard deviation is high, at 3.7%.

The provider level variation also accounts for a high percentage of total variation, at 1.2%. This means that relative to other indicators, a higher percentage of the variation occurs at the provider level, rather than the discharge level, though some of the variance remains at the discharge level. The signal ratio is only moderate, at 62.9%. This means that it is likely that some of the observed differences in provider performance do not represent true differences in provider performance. The high R-square (78.1%) reflects the higher proportion of signal that can be extracted using multivariate techniques.

**Bias.** Signal variance decreases by more than 25% with risk adjustment, suggesting that some of the observed variance is due to differences in patient characteristics. The indicator performs fairly on the multiple measures of minimum bias. The rank correlation is fair at 0.596. The impact on the extremes is large. Only 30.6% of providers in the highest decile remain, and only 60.3% in the lowest decile remain, after risk adjustment. Similarly, the number of providers moving at least two deciles in relative rank is also high. The absolute magnitude of risk adjustment is also substantial.

**Construct validity.** Pneumonia mortality loads on factor 1. It is positively related to mortality indicators, such as stroke, GI hemorrhage, and congestive heart failure, and to a lesser extent, hip fracture.

***Discussion***

Pneumonia is a leading cause of death in the United States. Treatment with appropriate antibiotics may reduce mortality. Hospitals with high mortality rates appear to have underlying quality problems when examining processes of care.

Though pneumonia mortality is common, patients are highly heterogeneous, potentially reducing the precision with which the indicator can be measured. Our empirical analysis found that the precision for this indicator was high, with high systematic variation and a moderate signal ratio. Multivariate techniques improved the amount of signal that can be extracted for this indicator. Thus, MSX smoothed estimates are recommended for this indicator to aid with precision problems due to random noise.

Our empirical analysis found substantial bias for this indicator, and it is likely that without risk adjustment, providers may be mislabeled as poor quality. The literature review notes that generic risk adjustment models derived from administrative data usually do not perform well with this indicator. Nonetheless, the potential for admitting more less severely ill patients, and thereby artificially deflating the mortality rate revisits the potential need for risk adjustment. Providers with particularly high and low mortality rates should examine the case-mix of their patients for comorbidities, age and clinical characteristics to test for simple or complex case-mixes. Chart reviews may be helpful in determining whether differences truly arise from quality of care, or patient level differences in coding, comorbidities, or severity of disease. Another source of bias is the outpatient management of some patients. Some providers may admit only the most severely ill patients and handle other patients in an outpatient setting, while others do not do this. This results in a more severe casemix. Providers may wish to examine rates of outpatient care.

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

Overall, this indicator is recommended for inclusion in the HCUP II QI set. It received an empirical rating of 7 out of 26, and smoothing is highly recommended. This indicator is

recommended with three major caveats of use. First, some pneumonia care occurs in an outpatient setting, and selection bias may be a problem for this indicator. Second, thirty-day mortality may be somewhat different than in-hospital mortality, leading to information bias. Third, risk adjustment for clinical factors, or at minimum APR-DRGs, is recommended due to the confounding bias for this indicator.

**INDICATOR 38: ACUTE STROKE MORTALITY RATE**

| | |
|---|---|
| Indicator | Provider level mortality rate for stroke. |
| Relationship to Quality | Better processes of care may reduce short-term mortality for stroke . As such lower rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Number of deaths per 100 discharges with principal diagnosis code of stroke. |
| Outcome of Interest | Number of deaths with principal diagnosis code for stroke (see Appendix 6). |
| Population at Risk | All discharges with principal diagnosis code for stroke (see Appendix 6).<br><br>Age 18 years and older.<br><br>Exclude transfers to other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates |

***Evidence from the literature***

**Face validity**. Stroke remains the third leading cause of death in the U.S. [756] Based on two population studies,[757, 758] Broderick et al conservatively estimates that approximately 725,000 acute strokes occur each year in the U.S..[758] Analysis of the 1995 HCUP database indicates that approximately 700,000 hospitalizations each year represent acute stroke. [176] However, hospital care has a relatively modest impact on patient survival, and most stroke deaths occur after the initial acute hospitalization.

**Precision**. Although strokes are very common, only 10-15% of stroke patients die during hospitalization. [176, 757, 759] Because stroke severity has a large effect on acute mortality, hospital mortality rates may be subject to considerable random variation.

**Minimum bias.** Four studies have shown significant inaccuracies in ICD-9 codes for identifying stroke patients.[758, 760-763] A previous study using the HCUP database[176] pooled these results to estimate the positive predictive value (PPV) of primary or secondary ICD-9 codes 430-438, which might potentially be used to identify stroke patients. The authors' estimates suggested that code 431 performed relatively well, although 14% of patients with this code will not have had an acute stroke. Moreover, this code only accounts for a minority of all strokes. Overall, codes 431, 434, and 436 probably provide the best combination of sensitivity and specificity. However, approximately 20% of patients with these codes actually have non-stroke diagnoses, and some strokes will be missed. Many of these patients will have undergone procedures related to stroke, such as cerebral angiography of carotid endarterectomy.[762] These groups of patients with different mortality risks than those with acute stroke. Although such "false positive" patients might be excluded based on having one of these procedures on the admission date (highly unlikely, especially for carotid endarterectomy), we found no studies that

investigated further steps to improve the predictive power of the discharge information. We also found no studies documenting cross-hospital variations in these coding practices, but significant variation seems possible.

The relative lack of effective medical therapy for acute stroke in the past led to considerable variation in hospital admitting practices. Geographic variations in admitting practice exert a significant effect on hospital mortality rates, far more than do variations in case-fatality rates.[764] Although it is possible that recent evidence on stroke treatments may have altered these historic patterns of variation, it seems likely that they remain large. Thus, comparisons within geographic areas are likely to be more valid than cross-area comparisons. In addition, because of the increased interest in the care of acute stroke patients, more patients with transient ischemic attacks (TIA) are likely to be admitted at some hospitals, whereas previously many such patients were considered safely managed as outpatients.[765] Because patients experiencing TIAs are much less likely to die in hospital than patients suffering acute stroke, hospitals with more liberal admitting policies may appear to have lower mortality rates, if these "rule-out" stroke patients are incorrectly coded as true strokes. Little evidence exists on the extent to which such coding anomalies exist.

Although strokes are associated with significant one-year mortality, early deaths tend to occur only in a subset of patients with the severest strokes (e.g., those in a coma at presentation). The majority of stroke deaths reflect medical complications in the weeks to months following admission.[757, 759, 766] Thus, hospitals with longer LOS, especially those that keep patients in hospital for rehabilitation therapy rather than discharging them to subacute care, may appear to have higher mortality rates. Investigating hospital LOS and transfer rates to rehabilitation hospitals in conjunction with stroke mortality may provide some evidence on this question.

*Risk adjustment*. Studies of the impact of risk adjustment based on discharge data suggest that a large number of hospitals identified as outliers using one system will have acceptable mortality rates using others.[10, 11] However, as noted above, mortality rates are a relatively noisy measure of hospital quality. Thus, it is possible that the different comorbidity systems were simply picking up the fact that the "true" quality of a hospital is not very well identified using simple adjusted stroke mortality rates. A recent prognostic model combined clinical and imaging variables to predict morbidity and mortality after acute stroke with excellent Receiver Operating Curve (ROC) characteristics.[767] Importantly, this model did not include "Do Not Resuscitate" orders as a predictor, a factor that detracted from the plausibility of previous predictions mortality models for stroke.[768] Most of the seven predictor variables in this recent model[767] are not reliably captured in current hospital discharge databases.

Some key mortality predictors might be captured. Coma at presentation confers a markedly increased risk death in acute stroke.[10] Use of mechanical ventilation on the first day of hospitalization quite reliably identifies stroke patients with coma at presentation.[769] Thus, if reliably measured, risk adjustment based on the use of mechanical ventilation on the first day of admission is a potentially important and feasible risk adjuster. A history of previous stroke substantially increases the mortality of patients admitted with stroke,[766] but non-longitudinal discharge data will not be able to distinguish first strokes from recurrent strokes.

Despite the smaller effect of aspirin on both primary and secondary prevention of stroke, patients with acute stroke who already are taking aspirin (e.g., for any other indication) tend to have better outcomes.[770, 771] Given these recent results, prior aspirin use might be an important omitted variable. However, no data exist on the extent to which prior aspirin use differs systematically across stroke patients admitted to different hospitals.

**Construct validity**. We located no articles specifically addressing construct validity at a hospital level. On a patient level processes of care have been shown to decrease mortality. However, it is unknown how implementing these processes of care would actually affect provider-level mortality rates. The publicity surrounding one randomized trial showing a benefit for thrombolytic therapy in acute stroke[772] generated considerable attention as ushering in a new era of stroke therapy. Importantly, though, the patients treated with thrombolytic therapy in all of the randomized trials represent only a small percentage of all stroke patients.[773] To derive benefit, patients must receive thrombolytic therapy within 3 hours of symptom onset,[774] but the majority of patients do not even arrive at the hospital within three hours.[775-777] Even among experienced neurologists and radiologists, considerable inter-rater agreement is only moderate for characterizing computed tomography findings that qualify patients for thrombolytic therapy.[778]

A recent study involving 57 U.S. medical centers managed to enroll only 389 patients over an almost 2 year study period,[779] and 13% of these patients in fact received thrombolysis after the 3 hour time window stipulated in the study protocol. Another recent study reported protocol violations (compared to national guidelines) in approximately 50% of patients, and a higher mortality overall for patients who received thrombolytic therapy.[780] Other studies have replicated the reported benefits of thrombolysis for acute stroke, but the existence of these conflicting results and the small percentage of stroke patients who receive thrombolysis suggest that more or less effective use of this treatment is likely to have only a modest impact on hospital mortality.

Other treatments have also shown only limited benefits for acute mortality. In contrast to the situation with acute myocardial infarction, the clinical benefit of aspirin appears modest.[781-784] Two large studies have demonstrated a significant benefit for aspirin in acute stroke,[785, 786] but again in contrast to acute myocardial infarction, little evidence exists on whether aspirin is underused in stroke management. Moreover, the magnitude of the impact of aspirin on acute stroke mortality is much smaller.[785-787]

In the RAND PPS Quality of Care study in 1990, patients with higher stroke process scale scores demonstrated significantly lower risk-adjusted 30-day stroke mortality on three out of four measured process subscales and on an overall process scale.[676]

The use of stroke units to care for patients with acute stroke (as opposed to caring for them on general medical wards) likely lowers mortality and improves functional outcomes.[788, 789] Overviews of the literature on stroke units reveal that the impact of stroke units has mostly been demonstrated in Europe.[788, 790] Evidence on their impact in the United States is more limited. Taken together, these factors suggest that inpatient care is unlikely to have a large impact on stroke mortality rates.

Stroke admission rates and mortality might be a more valid indicator of the quality of preventive ambulatory care. The underuse of anticoagulation to prevent embolic strokes in patients with atrial fibrillation does (in contrast to aspirin) represent an important quality problem.[791] Underuse of coumadin represents a failure of previous care, not the hospitalization for the acute stroke. Similarly, inadequate detection and control of hypertension, hypercholesterolemia, and other risk factors may lead to more frequent and more severe stroke admissions. One difficulty with using stroke outcomes for evaluating ambulatory care is that the coding of stroke subtypes (hypertensive, embolic, etc.) in discharge data is not very reliable.[762]

**Fosters true quality improvement**. The relatively low early compared to late mortality raises the possibility of a perverse incentive for institutions to lower their stroke mortality by

discharging patients to die in nursing homes. Thus, investigating rates of discharge to nonacute facilities or transfers to other hospitals might be important to consider in conjunction with in-hospital stroke mortality. "Overcoding" transient ischemic attacks as strokes may also result in decreasing stroke mortality rates; although in principle this should not occur, we found little evidence on whether such coding errors were frequent.

All in-hospital mortality measures may create perverse incentives to reduce hospital mortality by discharging patients earlier, and thereby shifting deaths to skilled nursing facilities or outpatient settings. This phenomenon may also lead to biased comparisons among hospitals with different mean lengths of stay. Although we found no data on the sensitivity of standardized stroke mortality measures (at the hospital level) to the inclusion or exclusion of post-discharge deaths, inpatient mortality among Medicare patients in California was only 76% of 30-day mortality in 1985.[106] This gap has presumably expanded over the last 15 years. This finding suggests that changes in length of stay may modestly alter the ranking of hospital performance using this measure.

**Prior use.** Stroke mortality has been used in the literature to a limited extent, and evaluations have focused on the limitations of the measure described above. [10, 63]

Several organizations have used stroke mortality indicators, including the University Hospital Consortium,[370] HealthGrades.com,[377] Maryland Hospital Association Quality Indicator Project,[369] and the Greater New York Hospital Association.[372] The Cleveland Health Quality Choice[374] project also includes a stroke mortality measure, however this indicator is risk adjusted using detailed clinical data. The Michigan Hospital Association[373] includes stroke mortality in its aggregate measure of in-hospital mortality.

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw provider level rate/standard deviation | 21.3%, 13.7% | |
| Systematic provider -level standard deviation** | 5.3% | High |
| Provider variation as a percentage of total variation** | 1.7% | High |
| Signal ratio** | 51.9% | Moderate |
| R-Square** | 70.7% | High |
| **APR-DRG, age-, gender- adjusted | | |
| *Minimum Bias – APR-DRG risk adjustment* | | |
| Signal variance change with risk adjustment | Decreases | Fair |
| Absolute impact: | | |
| Average absolute change (in %) | 13.1% | Good |
| Relative impact: | | |
| Rank correlation | 0.803 | Good |
| Percent remaining in high decile/low decile | 63.8% / 62.9% | Good/ Fair |
| Percent changing more than 2 deciles | 24.4% | Fair |

**Precision.** This indicator is precise, with a raw provider level mean of 21.3% and a standard deviation of 13.7%. The systematic provider level standard deviation is high, at 5.3%. The provider level variation also accounts for a high percentage of total variation, at 1.7%. This means that relative to other indicators, a higher percentage of the variation occurs at the provider level, rather than the discharge level, though some of the variance remains at the discharge level. The signal ratio is only moderate, at 51.9%. This means that it is likely that some of the

observed differences in provider performance do not represent true differences in provider performance. The high R-square (70.7%) reflects the higher proportion of signal that can be extracted using multivariate techniques.

**Bias.** Signal variance decreases substantially with risk adjustment, indicating that some of the true variation among providers reflects differences in patient characteristics. The indicator performs fairly to well on the measures of minimum bias. The rank correlation is good at 0.803, and risk adjustment impacts the lowest decile disproportionately to the highest, relative to other indicators. Further, 24.4% of providers change more than 2 deciles with risk adjustment. The absolute impact of risk adjustment is moderate.

**Construct validity.** Stroke mortality loads highly on factor 1. It is positively related to mortality indicators, such as pneumonia, GI hemorrhage, and congestive heart failure, and to a lesser extent, hip fracture.

***Discussion***

Acute stroke mortality has been the subject of study as a potential indicator. Our literature review noted some potential problems with the indicator, stemming from this literature. Quality treatment must be timely and efficient to prevent potentially fatal brain tissue death. Patients may not present until after the fragile window of time has past. Further, many deaths occur out of the hospital, suggesting that linkage to death records for patients post-discharge may be a good addition to this indicator.

This indicator is measured with adequate precision, having high provider systematic variation and a moderate signal ratio. Some of the observed variance may not in fact be true differences in performance. Mortality for stroke may be a rare event at small hospitals, reducing the precision for these providers. Multivariate techniques improve the amount of signal that can be extracted for this indicator. As such, MSX smoothing of data is recommended for this indicator, to help reduce precision problems due to random noise.

Our empirical analyses found some bias, when adjusting for age, sex and APR-DRGs. Risk adjustment is recommended for this indicator. Other risk adjustment systems have been reported to change relative provider performance for this indicator. Clinical factors of severity upon presentation, including mechanical ventilation on the first day, may vary systematically by hospital and do influence mortality. Providers with high rates may wish to examine the case-mix for these potentially complicating factors. Further, hospitals with rehabilitation programs may have higher rates than hospitals that discharge to sub-acute care. Providers may want to use this indicator in conjunction with length of stay for their hospitals and for surrounding areas. Coding appears suboptimal for acute stroke and may lead to bias. As such, some caution in interpreting this indicator is warranted.

Overall, this indicator is recommended for inclusion in the HCUP II QI set. It received an empirical rating of 10 out of 26, and smoothing is highly recommended. This indicator is recommended with three major caveats of use. First, some stroke care occurs in an outpatient setting, and selection bias may be a problem for this indicator. Second, thirty-day mortality may be somewhat different than in-hospital mortality, leading to information bias. Third, risk adjustment for clinical factors, or at minimum APR-DRGs, is recommended due to the confounding bias for this indicator.

## 3.E.6. Post-Procedural Mortality Measures

**INDICATOR 39: ABDOMINAL AORTIC ANEURYSM (AAA) REPAIR MORTALITY RATE**

| | |
|---|---|
| Indicator | Provider level mortality rate for AAA repair. |
| Relationship to Quality | Better processes of care may reduce mortality for AAA repair. As such lower rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Number of deaths per 100 discharges with procedure code of AAA repair. |
| Outcome of Interest | Number of deaths with procedure code for AAA repair (see Appendix 6). |
| Population at Risk | All discharges with procedure code of AAA repair and diagnosis of AAA (see Appendix 6) in any field.<br><br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |

***Evidence from the literature***

**Face validity**. Abdominal aortic aneurysm repair requires technical proficiency with the use of complex equipment. Technical errors may lead to clinically significant complications, such as arrhythmias, acute myocardial infarction, colonic ischemia, and death. Mortality is relatively high, especially if the aneurysm has already ruptured. Recent studies using North American population-based data sets have reported 3.5-6.2% in-hospital mortality after elective repair of intact aneurysms [188, 196, 294-296, 300, 301] and 40-55% in-hospital mortality after emergent repair of ruptured aneurysms.[62, 196, 299-301] These data suggest that improved quality of care could have a substantial impact on public health, despite the fact that the condition is relatively uncommon.

**Precision**. Abdominal aortic aneurysmectomy is not as common as the other cardiovascular procedures described in this report; only about 48,600 were performed in the USA in 1997 (1.8 per 10,000 persons).[293] Based on state all-payer databases, the mean annual frequency of abdominal aortic aneurysmectomies was 16.4-18.3 per hospital in Florida (unruptured only) in 1992-1996,[294] 8.4 per hospital in New York (unruptured only) in 1990-1995,[295] and 13.8 per hospital in Maryland in 1990-1995.[296] The relatively small number of abdominal aortic aneurysm resections performed by each hospital suggests that mortality rates are likely to be unreliable at the hospital level.

**Minimum bias**. Any measure based on an elective procedure, rather than a patient diagnosis, holds the potential for selection bias caused by the decision to elect surgery. Theoretically, one could account for the patient characteristics that are selected upon, but it is unlikely that administrative data are rich enough to do this comprehensively. Most previous studies of mortality following abdominal aortic aneurysmectomy have used administrative data,

so these studies have been unable to identify residual bias. The known predictors of in-hospital mortality include whether the aneurysm is intact or ruptured (see above), age, female gender, admission through an emergency room,[188, 294-296] various comorbidities such as renal failure and dysrhythmias,[195, 300] and Charlson's comorbidity index.[301] The largest study involving clinical data identified several additional predictors, the first three of which were more powerful than age: American Society of Anesthesiologists' risk class, leukocyte count, blood urea nitrogen, weight loss, and serum albumin.[298] A history of myocardial infarction and the number of myocardial segments with reversible thallium defects were independent predictors of major cardiac events after abdominal aortic aneurysmectomy in one study,[792] but the latter variable was non-significant in another study.[793] In the absence of studies explicitly comparing models with and without these additional clinical data elements, it is difficult to assess whether administrative data contain sufficient information to remove bias.

**Construct validity**. Several lines of evidence support the construct validity of this indicator. First, numerous studies (summarized above) have reported an association between hospital volume and mortality following abdominal aortic aneurysmectomy. The consistent association between volume and risk-adjusted mortality supports the validity of both measures of performance, and is consistent with the hypothesis that more experience leads to improved technical skills and better outcomes.

Second, previous studies have identified several other hospital or surgeon characteristics that are associated with lower mortality, after adjustment for patient characteristics ascertainable from administrative data. These characteristics include surgeon volume,[294, 296] especially for ruptured aneurysms,[62, 195, 299] board certification as a vascular surgeon,[294, 299] and having daily rounds by an intensive care physician.[197] Although two studies failed to show significant effects of surgeon volume,[195, 197] the correlation between hospital/physician characteristics and in-hospital mortality in most studies supports the validity of in-hospital mortality as a measure of quality.

Third, excessive blood loss was identified in one study as the most important predictor of mortality after elective abdominal aortic aneurysmectomy.[794] Excessive blood loss is a potentially preventable complication of surgery. Similarly, perioperative beta adrenergic blockade may help prevent cardiac deaths and nonfatal myocardial infarctions among high-risk patients undergoing major peripheral vascular surgery, including abdominal aortic aneurysmectomy.[795]

**Fosters true quality improvement**. One possible adverse effect of in-hospital mortality measures is to encourage earlier postoperative discharge. We are aware of no data on the likelihood or consequences of premature discharge after abdominal aortic aneurysmectomy. Another potential response would be to avoid operating on high-risk patients, although it is unclear to what extent providers could actually recognize and avoid high-risk patients. Indeed, many high-risk patients would actually benefit from being transferred to a more experienced center.

All in-hospital mortality measures may create perverse incentives to reduce hospital mortality by discharging patients earlier, and thereby shifting deaths to skilled nursing facilities or outpatient settings. This phenomenon may also lead to biased comparisons among hospitals with different mean lengths of stay. We found no published evidence about whether the difference between inpatient and 30-day mortality for AAA mortality is substantial enough to cause concern.

**Prior use.** Abdominal aortic aneurysmectomy mortality has not been widely used as an indicator of quality. It is used by HealthGrades.com.[377] Pennsylvania Health Care Cost Containment Council includes AAA repair in their "Other major vessel operations except heart (DRG 110)" indicator, though the indicator is defined using clinical risk adjustment.[681]

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw provider level rate/standard deviation | 21.5%, 26.8% | |
| Systematic provider -level standard deviation* | 6.7% | High |
| Provider variation as a percentage of total variation* | 3.5% | High |
| Signal ratio* | 30.7% | Low |
| R-Square* | 57.1% | Moderate |
| * age-, gender- adjusted | | |
| *Minimum Bias – APR-DRG risk adjustment* | | |
| Signal variance change with risk adjustment | Decreases | Fair |
| Absolute impact: | | |
| Average absolute change (in %) | 23.4% | Fair |
| Relative impact: | | |
| Rank correlation | 0.890 | Good |
| Percent remaining in high decile/low decile | 70.8% / 64.6% | Good/ Fair |
| Percent changing more than 2 deciles | 14.1% | Good |

**Precision.** This indicator is precise, with a raw provider level mean of 21.5% and a substantial standard deviation of 26.8%. The systematic provider level standard deviation is high, at 6.7%. The provider level variation also accounts for a high percentage of total variation, at 3.5%. This means that relative to other indicators, a higher percentage of the variation occurs at the provider level, rather than the discharge level. Finally, the signal ratio is low, at 30.7%. This means that it is likely that the some of the observed differences in provider performance do not represent true differences in provider performance. Additional signal can be extracted using multivariate techniques, as reflected by the moderate R-square. However, this signal remains lower than for other indicators.

**Bias.** Signal variance decreases by over 25% with risk adjustment, indicating that some of the true variation among providers reflects differences in patient characteristics. Risk adjustment for APR-DRGs was not available for this indicator, due to its distribution. Therefore, only age and sex risk adjustment was performed. The indicator performs fairly to well on the multiple measures of minimum bias. The rank correlation is high at 0.890, and both the high and low decile are impacted by risk adjustment. The percentage of providers moving at least two deciles in relative rank is moderate (14.1%). The absolute magnitude of the impact of risk adjustment is high.

**Construct validity.** AAA repair does not load substantially on any of the three retained factors. However, it does appear to be positively related to other post-procedural mortality measures, such as craniotomy ($r=.28$, $p<.0001$) and CABG ($r=.17$, $p<.01$).

### *Discussion*

AAA repair is technically difficult procedure, with a relatively high mortality rate. The main evidence for the validity of this indicator arises from the volume-outcome literature. Higher

volume hospitals have been noted to have lower mortality rates, suggesting some difference in the processes of care between lower and higher volume hospitals, resulting in better outcomes. What those processes are, if they truly exist, is not known.

AAA repair is an infrequent procedure, calling into question the precision of this indicator. Our empirical analyses confirmed this suspicion, though smoothing the indicator appeared to help tremendously, and is recommended for this indicator. After smoothing, this indicator is precise, with substantial systematic variation. However, the R-square is only moderate, suggesting that some of the observed variation does not reflect true differences in performance.

Little literature has been published on the potential bias of this indicator. It is known that comorbidities and clinical factors such as anesthesia risk and laboratory results do influence mortality rates. Due to the distribution of this indicator, we were not able to adjust using APR-DRGs. Our empirical analyses of demographic risk adjustment noted some potential bias for this indicator. Use of smoothed data (via MSX or other procedures) may help to avoid the erroneous labeling of a hospital as an outlier. Additional medical chart review and/or analyses of laboratory data may be helpful in determining whether more detail risk adjustment is necessary.

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

Overall, this indicator is recommended for inclusion in the HCUP II QI set. It received an empirical rating of 8 out of 26, and smoothing is highly recommended. This indicator is recommended with two major caveats of use. First, risk adjustment for clinical factors, is recommended due to the confounding bias for this indicator. Second, little evidence exists supporting the construct validity of this indicator.

**INDICATOR 40: CORONARY ARTERY BYPASS GRAFT (CABG) MORTALITY RATE**

| | |
|---|---|
| Indicator | Provider level mortality rate for CABG. |
| Relationship to Quality | Better processes of care may reduce mortality for CABG. As such lower rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Number of deaths per 100 discharges with procedure code of CABG. |
| Outcome of Interest | Number of deaths with diagnosis code for CABG (see Appendix 6). |
| Population at Risk | All discharges with ICD-9-CM code 36.1 in any procedure field. (see Appendix 6)<br><br>Age 40 years and older.<br><br>Exclude transfers to other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |

***Evidence from the literature***

**Face validity**. Post-CABG mortality rates have recently become the focus of several state public reporting initiatives.[40, 796, 797] Although the effect of the reporting on reducing post-CABG mortality and influencing cardiologists referral patterns may not be substantial[796, 797] or may even have had unintended consequences,[162] studies suggest that the reports are widely read and serve as the basis for discussions between physicians and patients about the risks associated with cardiac surgery.

**Precision**. One study that specifically considered the precision of the post-CABG mortality reports found that there is a significant amount of random variation, particularly for smaller hospitals due to reduced sample sizes.[40] By applying hierarchical statistical models intended to remove the random noise, the authors were able to detect some outliers that appeared to differ in performance not just because of chance. Without applying such sophisticated statistical methods, it is likely that hospitals will be identified as outliers as a result of random variations (patient and other factors beyond the hospital's control).

**Minimum bias**. Any measure based on an elective procedure, rather than a patient diagnosis, holds the potential for selection bias caused by the decision to elect surgery.[40, 162] Theoretically, one could account for the patient characteristics that are selected upon, but it is unlikely that administrative data are rich enough to do this comprehensively. On the basis of numerous studies using large databases from New York's Cardiac Surgery Reporting System,[798] the Northern New England Cardiovascular Disease Study Group,[799] the Cleveland Clinic,[800] and the Society of Thoracic Surgeons (among others),[801] there is general consensus that cardiac function, coronary disease severity, and the urgency of surgery are powerful predictors of mortality.[101] Hannan and colleagues demonstrated only moderate correlations (0.69-0.75) in

hospital risk-adjusted mortality rates between a model based on detailed clinical data and models based on either Medicare claims[799] or hospital discharge abstracts[799]; classification of hospital outliers differed substantially. These differences were largely attributable to two or three risk factors not available from administrative data. In another study, comparisons of hospital performance were relatively insensitive to the severity adjustment method, with correlations greater than 0.9[204] and code-based measures (e.g., APR-DRGs) had better statistical performance than two measures based on physiologic data.[17] These findings may be misleading, because the code-based measures included complications of care as well as comorbidities, and because the physiologic measures were not designed in accord with professional consensus.[799]

**Construct validity**. Several lines of evidence provide limited support for the construct validity of this indicator. First, numerous studies (summarized above) have reported an association between hospital volume and mortality following coronary bypass surgery. Sowden et al.[319] systematically reviewed 15 studies of the volume-outcome relationship for CABG; six used non-overlapping data and reported effect estimates for fixed volume categories. Among these six studies, the apparent benefit of high CABG volume (>200 cases per year) diminished as casemix adjustment improved. Because casemix adjustment was generally more complete in more recent studies, the authors could not exclude the possibility that the benefit of high volume actually decreased between 1972 and 1991. In addition, analyses using instrumental variables suggest that much, if not all, of the volume effect may be due to "selective referral" of patients to high-quality centers.[230, 239] These findings raise doubts about whether we can use the volume-outcome association as evidence of the construct validity of risk-adjusted mortality as a quality indicator. Studies of surgeon volume are similarly difficult to interpret, because of diminishing effects over time or counter-intuitive findings.[227, 233, 321]

The second source of evidence is that aortic crossclamp or perfusion time has been repeatedly associated with postoperative mortality, adjusting for a variety of patient characteristics.[802-804] In addition, longer cross-clamp times are associated with a higher incidence of postoperative atrial fibrillation.[805] Experienced surgeons and surgical teams should be able to reduce aortic crossclamp or perfusion time, thereby improving postoperative mortality. Mathew et al. also found that specific surgical techniques, including pulmonary vein venting and bicaval venous cannulation, may increase the risk of postoperative atrial fibrillation, which may in turn increase the risk of death. Perioperative use of beta blockers may have a cardioprotective effect, whereas use of nitrates may have a deleterious effect on in-hospital mortality.[806] Of course, patient-level reduction in mortality does not necessarily correspond with provider-level mortality. It is unknown how implementing these processes of care would actually affect provider-level mortality rates. Finally, several authors have reported on the experience of individual hospitals that responded to unfavorable risk-adjusted mortality data by identifying specific process failures. When these process failures were corrected, risk-adjusted mortality[160, 807] or length of stay[808] improved. None of these studies systematically compared processes of care between low-mortality and high-mortality hospitals, and all used risk-adjustment models that included physiologic predictors.

**Fosters true quality improvement**. One response physicians might make to public reporting of procedure-based mortality rates is to avoid operating on high-risk patients. Given the fact that high-risk patients may benefit the most from coronary bypass surgery, adverse selection is a serious concern. One study from the Cleveland Clinic reported a modest increase in the average annual volume of referrals from New York, from 61.4 before risk-adjusted mortality reports (1980-88) to 96.2 thereafter (1989-93). These patients were unusually high

risk, with an expected mortality rate 28% to 37% higher than that of Ohio residents.[160] In Pennsylvania, a drop from the 75th to the 25th percentile in the standardized, population-based rate of CABG surgery was associated with a 10% increase in the ratio of observed to expected mortality.[799] Nearly half (46%) of cardiothoracic surgeons surveyed in New York reported that one or more of their patients was "refused surgery last year, with the NYS CSRS (Cardiac Surgery Reporting System) being an integral part of the decision-making process."[809] Similarly, 63% of cardiothoracic surgeons surveyed in Pennsylvania reported that they were "less willing" to operate on the most severely ill patients since mortality data were first released; 59% of cardiologists reported that it had become "more difficult" to find surgeons willing to operate on such patients.[160] Yet according to Medicare data, the percentage of New York residents receiving bypass surgery out-of-state actually decreased in the early 1990's, with a concurrent increase (paralleling national trends) in the use of CABG after myocardial infarction.[810] Hence, there is no convincing evidence that providers are actually avoiding high-risk patients in states with public reporting. Although in-hospital mortality after CABG surgery decreased significantly in states that adopted outcome reporting (e.g., New York,[160] Northern New England[160]), in-hospital mortality after CABG surgery has also decreased in neighboring states despite the absence of statewide outcome reporting (e.g., Massachusetts).[160] A more recent analysis based on Medicare data found that risk-adjusted 30-day mortality after CABG surgery declined more rapidly in New York (10.3% per year) than in the rest of the nation (5.8% per year) between 1987 and 1992; New York and Northern New England had the lowest risk-adjusted CABG mortality rates in the US in 1992.[160] A better understanding of how and whether public reporting leads to improved quality, or what other factors may be at work, is required to understand the effect of ongoing statewide outcome studies.[204]

All in-hospital mortality measures may create perverse incentives to reduce hospital mortality by discharging patients earlier, and thereby shifting deaths to skilled nursing facilities or outpatient settings. This phenomenon may also lead to biased comparisons among hospitals with different mean lengths of stay. Although we found no data on the sensitivity of standardized CABG mortality measures (at the hospital level) to the inclusion or exclusion of post-discharge deaths, inpatient mortality among Medicare patients in California was over 90% of 30-day mortality in 1987-88.[811] Although this gap may have expanded over the last 15 years, the data suggest that changes in length of stay may have relatively little effect on the ranking of hospital performance using this measure.[812]

**Prior use.** Post-CABG mortality is publicly reported by several state health agencies, including the California CABG Mortality Reporting Project,[813] Pennsylvania Health Care Cost Containment Council,[681] New Jersey Department of Health and Senior Services,[799] and the New York Dept. of Health.[814] However, all of these programs are based on detailed clinical data systems. The Northern New England Cardiovascular Disease Study Group uses a similar data system, but reports risk-adjusted outcomes only to participating providers.[799] . Other recent users of CABG mortality rates as a quality indicator, with more limited risk-adjustment, include the University Hospital Consortium,[370] JCAHO's IMSystem, HealthGrades.com[377], Maryland Hospital Association (as part of the Maryland QI Project)[369], and Greater New York Hospital Association.[372]

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw provider level rate/standard deviation | 5.1%, 6.2% | |
| Systematic provider -level standard deviation** | 1.4% | Moderate |
| Provider variation as a percentage of total variation** | 0.5% | Moderate |
| Signal ratio** | 54.5% | Moderate |
| R-Square** | 69.8% | Moderate |
| **APR-DRG, age-, gender- adjusted | | |
| *Minimum Bias – APR-DRG risk adjustment* | | |
| Signal variance change with risk adjustment | Decreases | Fair |
| Absolute impact: | | |
| Average absolute change (in %) | 15.2% | Good |
| Relative impact: | | |
| Rank correlation | 0.743 | Fair |
| Percent remaining in high decile/low decile | 41.4% / 65.5% | Fair |
| Percent changing more than 2 deciles | 30.1% | Fair |

**Precision.** This indicator is precise, with a raw provider level mean of 5.1% and a standard deviation of 6.2%. The systematic provider level standard deviation is moderate, at 1.4%. The provider level variation also accounts for a moderate percentage of total variation, at 0.5%. This means that relative to other indicators, a lower percentage of the variation occurs at the provider level, rather than the discharge level. The signal ratio is only moderate, at 54.5%. This means that it is likely that the some of the observed differences in provider performance do not represent true differences in provider performance. The moderate R-square (69.8%) reflects the higher proportion of signal that can be extracted using multivariate techniques.

**Bias.** Signal variance decreases by more than 25% with risk adjustment, suggesting that some of the observed variance is due to differences in patient characteristics. The indicator performs fairly on the multiple measures of minimum bias. The rank correlation is fair at 0.743. The impact on the extremes is large. Only 41.4% of providers in the highest decile remain, and only 65.5% in the lowest decile remain, after risk adjustment. Similarly, the number of providers moving at least two deciles in relative rank is also high. However, the absolute magnitude of risk adjustment is moderate.

**Construct validity.** CABG mortality loads on factor 2. It is positively related to bilatateral catheterization, and negatively related to laparoscopic cholecystectomy.

### *Discussion*

CABG mortality is one of the most widely used and publicized post-procedural mortality indicators. However, we found limited evidence regarding the face or construct validity of this indicator.

The precision estimates for this indicator are somewhat lower than some of the other post-procedural indicators. However, CABG is a very common procedure increasing the importance of this indicator. The variance and precision are adequate for its use as a quality indicator. Multivariate techniques improves the ability to extract signal for this indicator, as such use of smoothed estimates will help avoid erroneous labeling of outlier hospitals, and is recommended.

This indicator is subject to substantial bias, as shown in our empirical analyses. Demographics, comorbidities, and clinical characteristics of severity of disease are important predictors of outcome that may vary systematically by provider. One study did note that APR-DRGs had the best statistical performance of any risk-adjustment system in predicting mortality, though this could be an artifact of including all codes that could be potential complications. Providers should risk adjust this indicator and may wish investigate potential case-mix differences using medical chart review when interpreting the results of this indicator. Chart review may also help distinguish comorbidities from complications, a potential pitfall of this indicator. Further, providers using APR-DRGs for risk adjustment may wish to screen for high rates of diagnoses that are potential complications of care, rather than comorbidities.

Hospital discharge practices differ, with some hospitals discharging patients earlier than others. For this reason, this indicator should be considered in conjunction with length of stay and transfer rates (though transfers are excluded in this indicator). While inpatient and 30-day mortality appear to be similar, providers may wish to track 30-day mortality if possible.

As with other mortality measures, there is concern that measuring mortality rates would result in access problems for higher risk patients. If possible, it may be useful to monitor the outcomes for CABG patients that do not undergo surgery. Chart review may also help to identify "best practices" to ensure real quality improvement.

Overall, this indicator is recommended for inclusion in the HCUP II QI set. It received an empirical rating of 5 out of 26, and smoothing is highly recommended. This indicator is recommended with four major caveats of use. First, as CABG is an elective procedure, some selection of the patient population may lead to bias. Further, providers may inflate the denominator by performing more CABG procedures on less clinically complex patients with questionable indications. Third, risk adjustment for clinical factors, or at minimum APR-DRGs, is recommended due to the confounding bias for this indicator. Finally, the evidence for the construct validity of this indicator is limited.

**INDICATOR 41: CRANIOTOMY MORTALITY RATE**

| | |
|---|---|
| Indicator | Provider level mortality rate for craniotomy. |
| Relationship to Quality | Better processes of care may reduce mortality for craniotomy. As such lower rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Number of deaths per 100 discharges with procedure code of craniotomy. |
| Outcome of Interest | Number of deaths with procedure code for craniotomy except for trauma (see Appendix 6). |
| Population at Risk | All discharges with procedure code for craniotomy, except for trauma (see Appendix 6).<br><br>Age 18 years or older.<br><br>Exclude transfers to other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates |

***Evidence from the literature***

**Face validity**. Craniotomy for the treatment of subarachnoid hemorrhage (SAH) and/or cerebral aneurysm is a condition and treatment that entails significant risk, with post-operative patient mortality and complications as the result of stroke, intracranial hypertension, systemic infections, hypoxia, pulmonary embolus, and cardiac arrythmias.[815] The procedure requires significant technical skill, and the ability to identify the most appropriate cases. Together with measures of volume and utilization, post-operative mortality rates will give a comprehensive perspective on provider performance for this condition.

**Precision**. Most providers perform relatively high number of procedures, with hospitals averaging 70 procedures per year for treatment of subarachnoid hemorrhage (SAH), and 30 per year for cerebral aneurysm.[816, 817] Post-operative mortality rates are also relatively high, averaging nearly 14% for patients 65+, [818] which will improve precision of the provider estimates.

**Minimum bias**. Studies have shown significantly higher post-craniotomy mortality rates by age group (from 3% for 23 to 39 year olds to 17% for 70+) for patients undergoing treatment for subarachnoid hemorrhage (SAH) [816, 819]. Other measures of health status and resource use, for example APACHE II score and ICU days, did not differ by age group, although these measures were generally higher for the elderly (65+) than the non-elderly (<65). For SAH, older patients generally present with more severe illness, including lower levels of consciousness, worse grade, a thicker subarachnoid clot , intraventricular hemorrhage, and hydrocephalus on admission. Older patients also present with higher comorbidity rates, including diabetes, hypertension, and pulmonary, myocardial, and cerebrovascular disease.[819] Age seems to have an independent effect on outcomes apart from these pre-existing conditions, suggesting

that the aging brain does not recover as well after initial bleeding. In summary, when controlling for age, a data element available on discharge abstracts, could also control for many other additional sources of bias across providers.

**Construct validity**. We located no evidence specifically evaluating the construct validity of this indicator. However, some evidence may lend to the validity of this indicator. Because the procedure risk is so high, provider skill may also be also be an important determinant in outcome.[817] Considering a post-operative mortality measure conjunction with a volume and utilization measure may offer the most comprehensive perspective on provider quality. Providers that perform more that 30 procedures per year have lower mortality than providers performing fewer than 30,[817, 818] although as we state in the introduction the volume-outcome relationship may be a product of patient selection. In one study, patients who were referred to a large medical center for SAH were less likely to have died early, and had fewer severe indications, including lower clinical grade, rate of coma, diastolic blood pressure, and younger patient age.[820] A utilization measure might help identify providers with greater propensity to perform the procedure.

**Fosters true quality improvement.** All in-hospital mortality measures may create perverse incentives to reduce hospital mortality by discharging patients earlier, and thereby shifting deaths to skilled nursing facilities or outpatient settings. This phenomenon may also lead to biased comparisons among hospitals with different mean lengths of stay. We found no published evidence about whether the difference between inpatient and 30-day mortality for craniotomy is substantial enough to cause concern.

**Prior use.** Post-operative mortality for craniotomy, non-trauma related is a measure used by the University Hospital Consortium (UHC).[370]

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw provider level rate/standard deviation | 16.2%, 18.5% | |
| Systematic provider -level standard deviation** | 3.7% | High |
| Provider variation as a percentage of total variation** | 1.5% | High |
| Signal ratio** | 28.9% | Low |
| R-Square** | 49.0% | Moderate |
| **APR-DRG, age-, gender- adjusted | | |
| *Minimum Bias – APR-DRG risk adjustment* | | |
| Signal variance change with risk adjustment | Decreases | Fair |
| Absolute impact: | | |
| Average absolute change (in %) | 33.8% | Fair |
| Relative impact: | | |
| Rank correlation | 0.786 | Good |
| Percent remaining in high decile/low decile | 29.6% / 61.1% | Fair |
| Percent changing more than 2 deciles | 27.4% | Fair |

**Precision.** This indicator is precise, with a raw provider level mean of 16.2% and a substantial standard deviation of 18.5%. The systematic provider level standard deviation is high, at 3.7%. The provider level variation also accounts for a high percentage of total variation, at 1.5%. This means that relative to other indicators, a higher percentage of the variation occurs at the provider level, rather than the discharge level, though some remains at the discharge level.

The signal ratio is low, at 28.9%. This means that it is very likely that the some of the observed differences in provider performance do not represent true differences in provider performance. The moderate R-square (49.0%) reflects the higher proportion of signal that can be extracted using multivariate techniques, though this remains lower than other indicators.

**Bias.** Signal variance decreases by more than 25% with risk adjustment. The indicator performs fairly on the multiple measures of minimum bias. The rank correlation is high at 0.786. Risk adjustment affects the extreme ends of the distribution substantially. Only 29.6% of providers in the highest decile remain after risk adjustment, and 61.1% in the lowest decile remain. Similarly, the number of providers moving at least two deciles in relative rank is also high.

**Construct validity.** Craniotomy does not load substantially on any of the three retained factors. However, it does appear to be positively related to other post-procedural mortality measures, such as AAA repair ($r=.28$, $p<.0001$) and CABG ($r=.23$, $p<.0001$), as well as stroke mortality ($r=.49$, $p<.0001$).

***Discussion***

Craniotomy is a complex procedure requiring surgical skill. Providers with high rates have better outcomes, though this may be an artifact of patient selection. We found little further evidence on the face validity or construct validity of this indicator.

This indicator is measured with good precision, with very high provider systematic variation. However, the signal ratio is low, suggesting that some of the observed variation may not reflect true differences in performance. Post-operative mortality rates are relatively high, and the surgery is relatively common for most providers. Such high provider variation suggests that improving quality of care could potentially greatly improve outcomes. Multivariate techniques improve the ability to extract signal for this indicator and are recommended. Use of smoothed estimates (via MSX or other methods) will aid in reducing precision problems due to random noise, such as the erroneous labeling of providers as “outliers.”

Using our empirical analyses we identified substantial bias for this indicator, particularly for age. Our literature review also suggests that age is an important factor to consider for risk adjustment. Older patients also have more severe illness and comorbidities, though these do not explain all of the higher mortality rates observed in an elderly population. Providers should risk adjust for age and comorbidities. Further examination of other patient characteristics that increase case-mix complexity may be completed through medical chart reviews or analyses of laboratory tests.

As with other mortality measures, there is concern that measuring mortality rates would result in access problems for higher risk patients. If possible, it may be useful to monitor the outcomes for subarachnoid hemorrhage or cerebral aneurysm patients that do not undergo surgery.

Overall, this indicator is recommended for inclusion in the HCUP II QI set. It received an empirical rating of 6 out of 26, and smoothing is highly recommended. This indicator is recommended with two major caveats of use. First, risk adjustment for clinical factors, or at minimum APR-DRGs, is recommended due to the confounding bias for this indicator. Second, little evidence exists supporting the construct validity of this indicator.

## INDICATOR 42: ESOPHAGEAL RESECTION MORTALITY RATE

| | |
|---|---|
| Indicator | Provider level mortality rate for esophageal resection. |
| Relationship to Quality | Better processes of care may reduce mortality for esophageal resection. As such lower rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Number of deaths per 100 patients with discharge procedure code of esophageal resection. |
| Outcome of Interest | Number of deaths with procedure code for esophageal resection (see Appendix 6). |
| Population at Risk | All discharges with procedure code of esophageal resection (see Appendix 6) and diagnosis code for esophageal cancer in any field.<br><br>Exclude transfers to other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |

***Evidence from the literature***

**Face validity.** Esophageal resection is a complex procedure that requires technical skill. The primary evidence for this indicator arises from the volume-outcome literature. Several studies have found that hospitals that perform more procedures have better mortality rates than lower volume hospitals. The magnitude of this relationship is relatively large as compared to other procedures. A full review of this literature can be found in the discussion of pancreatic resection as a volume indicator. This relationship suggests that there may be some differences in processes of care that result in better outcomes. Those processes have not been identified and are subject to controversy, as it is unclear what the causal relationship is, if there truly is one, between hospital volume and mortality.

**Precision.** Esophageal resection is a relatively uncommon procedure, which may impact the precision of the indicator. Patti et al[198] noted that most hospitals perform 10 or fewer procedures during a 5-year period. Utilizing several years of data, which has been done in some of the volume-outcome research, may help improve the precision of this indicator.

**Minimum bias.** Although we located no studies specifically addressing the need for risk adjustment, most of the volume-outcome studies published have used some sort of risk adjustment, suggesting that risk adjustment may be important for this procedure. Most of those studies used administrative data for risk adjustment.

**Construct validity.** Beyond the volume-outcome relationship we found no evidence for the construct validity of this procedure. Two studies have examined hospital volume as compared to in-hospital mortality rates. Patti et al.[198] used five volume categories, finding decreasing mortality rates of 17%, 19%, 10%, 16%, and 6% (1-5, 6-10, 11-20, 21-30, and >30 procedures during the 5-year study period). Gordon et al.[322] combined all complex

gastrointestinal procedures, finding that low volume (11-20 procedures per year) hospitals had an adjusted odds of death of 4.0 as compared to the one high volume hospital.

**Fosters true quality improvement.** Though we found no evidence on whether or not this indicator would stimulate true improvement in quality, it is possible that high risk patients may be denied surgery.

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

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw provider level rate/standard deviation | 20.2%, 36.6% | |
| Systematic provider -level standard deviation** | 2.4% | Moderate |
| Provider variation as a percentage of total variation** | 0.8% | Moderate |
| Signal ratio** | 8.9% | Low |
| R-Square** | 21.0% | Low |
| **age-, gender- adjusted | | |
| *Minimum Bias – APR-DRG risk adjustment* | | |
| Signal variance change with risk adjustment | | |
| Absolute impact: | | |
| Average absolute change (in %) | 16.8% | Good |
| Relative impact: | | |
| Rank correlation | 0.858 | Good |
| Percent remaining in high decile/low decile | 66.7% / 100% | Good/ V.G. |
| Percent changing more than 2 deciles | 9.5% | Good |

**Precision.** This indicator is precise, with a raw provider level mean of 20.2% and a substantial standard deviation of 36.6%. The systematic provider level standard deviation is moderate, at 2.4%. The provider level variation also accounts for a moderate percentage of total variation, at 0.8%. This means that relative to other indicators, a smaller percentage of the variation occurs at the provider level, rather than the discharge level. Finally, the signal ratio is low, at 8.9%. This means that it is very likely that the some of the observed differences in provider performance do not represent true differences in provider performance. Multivariate techniques improve the amount of signal that can be extracted, although the R-square is still low relative to other indicators.

**Bias.** Signal variance decreases by over 25% with risk adjustment, indicating that some of the true variation among providers reflects differences in patient characteristics. Due to the distribution of this indicator, APR-DRG risk adjustment was not available. Thus, only age and sex risk adjustment was performed. The indicator performs well on the multiple measures of minimum bias. The rank correlation is high at 0.858. Risk adjustment does seem to impact disproportionately at the extreme high end, as there no impact at the low end. The absolute impact of risk adjustment is moderate.

**Construct validity.** Since the distribution of this indicator violates the assumptions of factor analysis, this indicator was not included in our analysis of construct validity.

### *Discussion*

Esophageal resection is a complex cancer surgery. Several studies have noted that providers with higher volumes have lower mortality rates for the procedure than providers with

lower volumes. This suggests that perhaps providers with higher volumes have some characteristics, either structurally or with regard to processes, that influence mortality after this procedure. However, if these characteristics do indeed exist, what they are is unclear.

This indicator has moderate provider systematic variation, and the signal ratio is quite low. This indicates that some observed differences are not true differences in performance. Smoothing is somewhat helpful for this indicator and is recommended, though the amount of extractable signal remains low. As this procedure is performed only by a select number of hospitals, a majority of hospitals will have no cases in a year. The low numbers of these procedures on a provider level may compromise the precision of this indicator. Providers may wish to examine several consecutive years to potentially increase the precision of this indicator.

This indicator generally performed well on our tests of minimum bias, with some moderate bias identified. However, due to the distribution of this indicator only demographic risk adjustment was applied. All studies reviewed in the literature review risk adjusted the mortality rate in some manner, suggesting that risk adjustment is considered important. However, as the aims of the studies were to establish volume-outcome relationships, the extent to which adjusting affected provider performance was not reported. It is recommended that this indicator be risk adjusted.

Overall, this indicator is recommended for inclusion in the HCUP II QI set. It received an empirical rating of 8 out of 26, and smoothing is highly recommended. This indicator is recommended with two major caveats of use. First, risk adjustment for clinical factors, is recommended due to the confounding bias for this indicator. Second, little evidence exists supporting the construct validity of this indicator.

**INDICATOR 43: HIP REPLACEMENT MORTALITY RATE**

| | |
|---|---|
| Indicator | Provider level mortality rate for hip replacement. |
| Relationship to Quality | Better processes of care may reduce mortality for hip replacement. As such lower rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Number of deaths per 100 patients with discharge procedure code of partial or full hip replacement. |
| Outcome of Interest | Number of deaths with procedure code for hip replacement (see Appendix 6). |
| Population at Risk | All discharges with procedure code of partial or full hip replacement in any field (see Appendix 6).<br><br>Include only discharges with uncomplicated cases: diagnosis or procedure codes for osteoarthrosis of hip in any field (see Appendix 6).<br><br>Exclude transfers to other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |

***Evidence from the literature***

**Face validity**. Total hip arthroplasty (without hip fracture) is an elective procedure performed to improve function and relieve pain among patients with chronic osteoarthritis, rheumatoid arthritis, or other degenerative processes involving the hip joint. Mortality is very low, as it should be for a procedure that is designed to improve function rather than extend survival. However, patients who undergo total hip arthroplasty are often elderly, with multiple comorbidities. As a result, they are at significant risk of such postoperative complications as pneumonia, osteomyelitis, myocardial ischemia, and deep vein thrombosis. If these complications are not recognized and effectively treated, life-threatening problems such as sepsis, myocardial infarction, and pulmonary embolus may ensue. The ICD-9-CM definition of this procedure (81.51) is limited to primary arthroplasties; revisions are assigned a different code (81.53) that does not distinguish hemiarthroplasties from total arthoplasties.

**Precision**. Primary total hip arthroplasty is one of the most frequent types of major orthopedic surgery; about 160,000 were performed in the USA in 1998 (5.9 per 10,000 persons).[711] Based on state all-payer databases, the mean frequency of primary total hip arthroplasties was 72.8 per hospital in Ontario[821] and 106.9 per hospital in Florida (including total knee arthroplasties) in 1992.[234] However, the in-hospital or 30-day postoperative mortality rate in various studies ranged from 0.10% at New York's Hospital for Special Surgery[822] and 0.15% at Massachusetts General Hospital[823] to 1.97% (0.95% for indications other than hip fracture) in a 5% random sample of Medicare beneficiaries over 65 years of age.[824] The

relatively small number of deaths following total hip arthroplasty at each hospital suggests that mortality rates are likely to be unreliable at the hospital level. For example, age and sex standardized postoperative mortality varied 4.8 fold across 10 hospitals in the Oxford region, but this variation was not statistically significant.[825]

**Minimum bias**. Any measure based on an elective procedure, rather than a patient diagnosis, holds the potential for selection bias caused by the decision to elect surgery. Theoretically, one could account for the patient characteristics that are selected upon, but it is unlikely that administrative data are rich enough to do this comprehensively. The known predictors of in-hospital mortality include age, hip fracture, and the presence of any significant comorbidity.[716, 821, 824] Failure to adjust for hip fracture has been shown to introduce bias against certain hospitals in the Medicare Hospital Information Project.[826] Indication for surgery (other than hip fracture) and race have not been shown to predict postoperative mortality, whereas there is conflicting evidence on the effect of gender.[825] We are not aware of any studies that tested the effect of physiologic factors (available only from clinical data) on mortality. In the absence of studies explicitly comparing models with and without clinical data elements, it is difficult to assess whether administrative data contain sufficient information to remove bias.

**Construct validity**. We located no studies explicitly evaluating the construct validity of this indicator. However, there is limited evidence supporting an association between hospital volume and mortality following total hip arthroplasty. (Following Halm, Lee, and Chassin,[80] we did not find this evidence to be sufficiently strong to recommend total hip arthroplasty volume as a separate volume indicator.) Using administrative data from Florida, without any risk-adjustment, Lavernia and Guzman[234] found no association between hospital volume and mortality. However, surgeons with fewer than 10 cases per year showed a significant increase in the death rate, and hospitals with fewer than 10 cases per year showed a significant increase in the complication rate. In a similar analysis of Ontario data,[821] surgeon and hospital volumes were not significantly associated with mortality, postoperative infection, serious complications, or revision. A study of all Medicare claims from 1993 and 1994 showed lower in-hospital and in-hospital plus 30-day mortality after total hip arthroplasty at hospitals with higher Medicare volume for DRG 209 ("major joint and limb reattachment procedures, including primary and revision hip, knee, shoulder, and wrist arthroplasties").[236] By contrast, another study of Medicare data from 1979 and 1980 showed no association with 60-day mortality, after adjusting for age, sex, hip fracture, and medical school affiliation.[716] Older studies are also inconsistent. Maerki[718] and Luft[239] found a hospital volume effect using 1972 data from the Commission on Professional and Hospital Activities, whereas Farley[230] did not, using 1980-87 data from HCUP. These inconsistencies provide very limited support for the construct validity of mortality as a quality indicator.

More persuasive, perhaps, is mounting evidence that thromboembolic prophylaxis substantially reduces the incidence of symptomatic pulmonary embolism after elective total hip arthroplasty (e.g., 0.16% with warfarin, 0.26% with pneumatic compression, 0.36% with low molecular weight heparin, 1.51% with placebo).[716]

Although pulmonary embolism is known to be a major cause of death after hip arthroplasty,[827] there is still no clear evidence that thromboembolic prophylaxis (or any other specific process of care) reduces mortality after total hip arthroplasty. One observational study attributed a decrease in postoperative mortality from 0.36% in 1981-85 to 0.10% in 1987-91 to changes in perioperative care, such as reduced intraoperative blood loss, more aggressive arterial and oximetric monitoring, and increased use of epidural instead of general anesthesia.[822]

**Fosters true quality improvement**. One possible adverse effect of in-hospital mortality measures is to encourage earlier postoperative discharge. We are aware of no data on the likelihood or consequences of premature discharge after primary total hip arthroplasty. Another potential response would be to avoid operating on high-risk patients, although it is unclear to what extent providers could actually recognize and avoid high-risk patients. Indeed, many high-risk patients would actually benefit from being transferred to a more experienced center.

**Prior use.** In-hospital mortality following total hip arthroplasty is a current indicator in the HCUP I QI set. It is also reported by the HCFA in the Medicare Quality of Care Surveillance System,[828] and used by HealthGrades.com,[377] and the Greater New York Hospital Association.[372] Hip replacement is combined with all hip procedures in an indicator used by the Pennsylvania Health Care Cost Containment Council.[681] Finally, the University Hospital Consortium combines both knee replacement and hip replacement in a mortality indicator.[370]

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw provider level rate/standard deviation | 1.2%, 5.7% | |
| Systematic provider -level standard deviation** | 0.9% | Moderate |
| Provider variation as a percentage of total variation** | 1.2% | High |
| Signal ratio** | 20.0% | Low |
| R-Square** | 21.6% | Low |
| **APR-DRG, age-, gender- adjusted | | |
| *Minimum Bias – APR-DRG risk adjustment* | | |
| Signal variance change with risk adjustment | Decreases | Fair |
| Absolute impact: | | |
| Average absolute change (in %) | 48.9% | Fair |
| Relative impact: | | |
| Rank correlation | 0.642 | Fair |
| Percent remaining in high decile/low decile | 27.8% / 70.8% | Fair |
| Percent changing more than 2 deciles | 36.1% | Fair |

**Precision.** This indicator is adequately precise, with a raw provider level mean of 1.2% and a substantial standard deviation of 5.7%. The systematic provider level standard deviation is moderate, at 0.9%. The provider level variation also accounts for a high percentage of total variation, at 1.2%. This means that relative to other indicators, a high percentage of the variation occurs at the provider level, rather than the discharge level. The signal ratio is low, at 20.0%. This means that it is very likely that some the observed differences in provider performance do not represent true differences in provider performance. The R-square remains low, reflecting the minimal impact of multivariate signal extraction techniques.

**Bias.** Signal variance decreases by more than 25% with risk adjustment. The indicator performs fairly on the multiple measures of minimum bias. The rank correlation is fair at 0.642. Risk adjustment affects the extreme ends of the distribution substantially. Only 27.8% of providers in the highest decile remain after risk adjustment, and 70.8% in the lowest decile remain. Similarly, the number of providers moving at least two deciles in relative rank is also high. The average absolute change in performance relative to the mean is over 78%.

**Construct validity.** Hip replacement mortality is not strongly related to other indicators.

***Discussion***

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

The low mortality rate is likely to affect the precision for this indicator. Our empirical analyses confirmed that this indicator is measured with low precision. The signal ratio is especially low. It is likely that some of the precision seen does not reflect true differences in performance. Multivariate techniques do not improve the ability to extract signal for this indicator, but are recommended, as this indicator is very noisy. Nonetheless, this indicator has adequate precision for use as a quality indicator.

As hip replacement is an elective procedure, it is subject to selection bias. Patient characteristics such as age and comorbidities may influence the mortality rate for the procedure (particularly a diagnosis of hip fracture), and bias has been documented for this indicator. Our empirical analyses also identified substantial bias in this indicator, especially for providers at the extremes. This may result in the erroneous labeling of outlier providers. Risk adjustment is highly recommended for this indicator. Given the concerns raised in the literature, providers desiring to use this indicator may want to examine the case mix of their population.

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

Overall, this indicator is recommended for inclusion in the HCUP II QI set. It received an empirical rating of 3 out of 26, and smoothing is highly recommended. This indicator is recommended with three major caveats of use. First, as hip replacement is an elective procedure, some selection of patient population may create bias. Second, risk adjustment for clinical factors, or at minimum APR-DRGs, is recommended due to the confounding bias for this indicator. Finally, the evidence supporting the construct validity of this indicator is limited.

## INDICATOR 44: PANCREATIC RESECTION MORTALITY RATE

| | |
|---|---|
| Indicator | Provider level mortality rate for pancreatic resection. |
| Relationship to Quality | Better processes of care may reduce mortality for pancreatic resection. As such lower rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Number of deaths per 100 patients with discharge procedure code of pancreatic resection. |
| Outcome of Interest | Number of deaths with procedure code for pancreatic resection in any field (see Appendix 6). |
| Population at Risk | All discharges with procedure code of pancreatic resection (see Appendix 6) and diagnosis code for cancer in any field.<br><br>Exclude transfers to other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). |

***Evidence from the literature***

**Face validity.** Pancreatic resection is a complex procedure that requires technical skill. The primary evidence for this indicator arises from the volume-outcome literature. Several studies have found that hospitals that perform more procedures (which sometimes included other complex gastrointestinal procedures) have better mortality rates than lower volume hospitals. The magnitude of this relationship is relatively large as compared to other procedures. A full review of this literature can be found in the discussion of pancreatic resection as a volume indicator. This relationship suggests that there may be some differences in processes of care that result in better outcomes. Those processes have not been identified and are subject to controversy, as it is unclear what the causal relationship is, if there truly is one, between hospital volume and mortality.

**Precision.** Pancreatic resection is a relatively uncommon procedure, which may impact the precision of the indicator. Glasgow et al. [199] found that in California most hospitals perform 10 or fewer procedures during a 5 year period. However, the mortality rate is high, ranging from 4%-13%.[65, 238] Utilizing several years of data, which has been done in some of the volume-outcome research, may help improve the precision of this indicator.

**Minimum bias.** Although we located no studies specifically addressing the need for risk adjustment, most of the volume-outcome studies published have used some sort of risk adjustment, suggesting that risk adjustment may be important for this procedure. Most of those studies used administrative data for risk adjustment.

**Construct validity.** Beyond the volume-outcome relationship we found no evidence for the construct validity of this indicator. Ten studies have examined the volume relationship with in-hospital mortality. See the literature review for Pancreatic Resection Volume for details.

Glasgow and Mulvihill[199] estimated risk-adjusted, though the risk adjustment was limited, mortality rates of 14%, 10%, 9%, 7%, 8%, and 4% across six hospital volume categories (e.g., 1-5, 6-10, 11-20, 21-30, 31-50, and >50 procedures during the 5-year study period). Gordon et al.[322] estimated that the adjusted odds of death at minimal-volume (<11 "complex gastrointestinal procedures"/year) and low-volume (11-20 procedures/year) hospitals were 12.5 and 10.4 times that at a high-volume hospital (214 procedures/year). However, the generalizability of these results is limited by the fact that the last category included only one hospital.

Lieberman et al.[324] used 1984-91 hospital discharge data from New York State to analyze the association between mortality after pancreatic cancer resection and both physician and hospital volumes. The standardized mortality rate was 19%, 12%, 13%, and 6% at minimal (<10 patients during the 8-year study period), low (10-50 patients), medium (51-80 patients), and high-volume (>80 patients) hospitals, respectively. Surgeon volume was less significantly associated with mortality (6-13% risk-adjusted mortality across 3 volume categories); this effect disappeared in a model that included both physician and hospital volume, as confirmed by Sosa et al.[325] Studies using administrative data from Ontario[326], the United Kingdom,[327] and Medicare[328] have generated results similar to those from California and New York.

Gordon et al.[331] estimated that 61% of the observed reduction in statewide deaths among patients undergoing the Whipple procedure was attributable to the increasing market share of one facility, from 20.7% to 58.5% between 1984 and 1995.

**Fosters true quality improvement.** Though we found no evidence on whether or not this indicator would stimulate true improvement in quality, it is possible that high risk patients may be denied surgery.

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

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw provider level rate/standard deviation | 15.4%, 31.3% | |
| Systematic provider -level standard deviation** | 4.2% | High |
| Provider variation as a percentage of total variation** | 3.2% | High |
| Signal ratio** | 16.5% | Low |
| R-Square** | 34.7% | Low |
| ** age-, gender- adjusted | | |
| *Minimum Bias – APR-DRG risk adjustment* | | |
| Signal variance change with risk adjustment | Decreases | Fair |
| Absolute impact: | | |
| Average absolute change (in %) | 41.9% | Fair |
| Relative impact: | | |
| Rank correlation | 0.540 | Fair |
| Percent remaining in high decile/low decile | 71.4% / 28.6% | Good / Fair |
| Percent changing more than 2 deciles | 38.1% | Fair |

**Precision.** This indicator is moderately precise, with a raw provider level mean of 15.4% and a standard deviation of 31.3%. The systematic provider level standard deviation is high, at 4.2%. The provider level variation also accounts for a high percentage of total variation, at 3.2%. The signal ratio is low, at 16.5%. This means that it is very likely that the some of the

observed differences in provider performance do not represent true differences in provider performance. Multivariate extraction techniques do extract additional signal, although the R-square remains low relative to other indicators.

**Bias.** Signal variance decreases by over 25% with risk adjustment, indicating that some of the true variation among providers reflects differences in patient characteristics. Due to the distribution of this indicator, APR-DRG risk adjustment was not available. Thus, only age and sex risk adjustment was performed. The indicator performs fairly on the multiple measures of minimum bias. The rank correlation is low, indicating substantial impact of risk adjustment on relative performance. Risk adjustment affects the lowest decile disproportionately to the highest decile, with only 28.6% remaining in the lowest decile after risk adjustment, and 71.4% remaining in the highest decile. The absolute impact of risk adjustment is also large.

**Construct validity.** Since the distribution of this indicator violates the assumptions of factor analysis, this indicator was not included in our analysis of construct validity.

***Discussion***

Pancreatic resection is a complex cancer surgery. Several studies have noted that providers with higher volumes have lower mortality rates for the procedure than providers with lower volumes. This suggests that perhaps providers with higher volumes have some characteristics, either structurally or with regard to processes, that influence mortality after this procedure. However, if these characteristics do indeed exist, what they are is unclear.

This indicator has high provider systematic variation, though the signal ratio is quite low. This suggests that some of the observed variation does not reflect true differences in performance. Multivariate techniques do improve the amount of signal that can be extracted, and as a result smoothing is recommended. However, the amount of extractable signal remains lower than for other indicators. As this procedure is performed only by a select number of hospitals, a majority of hospitals will have no cases in a year. The low numbers of these procedures on a provider level may compromise the precision of this indicator. Providers may wish to examine several consecutive years to potentially increase the precision of this indicator.

This indicator generally performed poorly on our tests of minimum bias, suggesting that this indicator is subject to substantial bias. In addition, due to the distribution of this indicator, only age-sex risk adjustment was performed. All studies reviewed in the literature review risk adjusted the mortality rate in some manner, suggesting that risk adjustment is considered important. However, as the aims of the studies were to establish volume-outcome relationships, the extent to which adjusting affected provider performance was not reported.

Overall, this indicator is recommended for inclusion in the HCUP II QI set. It received an empirical rating of 5 out of 26, and smoothing is highly recommended. This indicator is recommended with two major caveats of use. First, risk adjustment for clinical factors, is recommended due to the confounding bias for this indicator. Second, little evidence exists supporting the construct validity of this indicator.

**INDICATOR 45: PEDIATRIC HEART SURGERY MORTALITY RATE**

| | |
|---|---|
| Indicator | Provider level mortality rate for pediatric heart surgery |
| Relationship to Quality | Better processes of care may reduce mortality for pediatric heart surgery. As such lower rates represent better quality care. |
| Benchmark | State, regional, or peer group average. |

Method:

| | |
|---|---|
| Quality Measure | Number of deaths per 100 patients with discharge procedure code of pediatric heart surgery. |
| Outcome of Interest | Number of deaths with diagnosis code for pediatric heart surgery (see Appendix 6). |
| Population at Risk | All discharges, age <18 years, with 1) procedure code of specified pediatric heart surgery in any field or 2) any heart surgery and a diagnosis for hypoplastic left heart syndrome (see Appendix 6).<br><br>See Appendix 6 for additional exclusions.<br>Exclude transfers to other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium). |

***Evidence from the literature***

**Face validity**. Pediatric cardiac surgery requires technical proficiency with the use of complex equipment. Technical errors may lead to clinically significant complications, such as arrhythmias, congestive heart failure, and death. It is thought that postoperative mortality rates vary considerably across hospitals, in a manner that reflects quality of care. Studying provider volume and mortality together would offer a comprehensive perspective on provider performance for pediatric cardiac surgery.

**Precision**. Previous studies suggest that pediatric cardiac surgery is highly concentrated at a relatively small number of facilities (e.g., 16 hospitals in New York, 37 in California and Massachusetts together). Although some of these facilities have very high volumes, a significant number (e.g., 16 hospitals in California and Massachusetts) perform fewer than 10 cases per year. The highly skewed volume distribution may have an adverse effect on the precision of this measure.

**Minimum bias**. Pediatric cardiac surgery represents a composite of numerous procedures performed to repair or palliate numerous congenital anomalies. The extreme heterogeneity among these procedures, and among the underlying anomalies, makes bias a serious concern. For example, among procedures with at least 100 cases in New York's Cardiac Surgery Reporting System[194] in 1992-1995, in-hospital mortality varied from 0.4% for repair of atrial septal defect (ASD) to 34.2% for Norwood repair of hypoplastic left ventricle. Even for a single procedure at major centers, such as the Fontan operation for tricuspid atresia or single ventricle, mortality depends heavily on physiological and functional factors, such as asplenia, atrioventricular valvular function, and mean pulmonary artery pressure .[829, 830] Technical

factors such as the dimension of the native pulmonary arteries may also be important.[831] Because these factors are not available in administrative data sets, and because the most complex patients are likely to be referred to selected centers, unmeasured risk factors could seriously confound inter-provider performance comparisons based on administrative data.

**Construct validity**. The evidence for the construct validity of this indicator comes from two sources. First, three studies (including one that used prospectively collected clinical data) have reported an association between hospital volume and mortality following pediatric cardiac surgery. Using a multivariate model that included age, complexity category, and four comorbidities, Hannan et al.[194] found 8.26% risk-adjusted mortality at hospitals with fewer than 100 cases per year, versus 5.95% at higher volume hospitals (an effect limited to surgeons who performed at least 75 cases per year). Two other studies using hospital discharge data from California and Massachusetts found similar effects of hospital volume .[295, 332] The consistent association between volume and risk-adjusted mortality supports the validity of both measures of performance, and is consistent with the hypothesis that more experience leads to improved technical skills and better outcomes. Other studies from single centers have confirmed this hypothesis by demonstrating improvements in mortality over time for a variety of procedures.[832-834]

The second source of evidence is that cardiopulmonary bypass or aortic crossclamp time has been repeatedly associated with postoperative mortality, adjusting for a variety of patient characteristics.[830, 835-837] This relationship has been demonstrated not just for the Fontan procedure, but also for the Norwood procedure for hypoplastic left heart syndrome.[838] Experienced surgeons and surgical teams should be able to reduce cardiopulmonary bypass or aortic crossclamp time, thereby improving postoperative mortality. It should be noted that patient-level reduction in mortality does not necessarily correspond with provider-level mortality. It is unknown how implementing these processes of care would actually affect provider-level mortality rates.

**Fosters true quality improvement**. One potential response by physicians to public reporting of procedure mortality rates would be to avoid operating on high-risk patients. Given that the risk factors for adverse outcomes after the more frequent procedures are well known to pediatric cardiac surgeons, and that many of these risk factors are not available from administrative data, avoidance of high-risk cases is a genuine concern. Although such behavior may lead to bias in estimating provider-specific performance, it would be unlikely to worsen population outcomes, because the indications for surgery are generally clear and many high-risk patients would actually benefit from being transferred to a more experienced center.

Another potential response by physicians to reporting in-hospital mortality would be to discharge patients earlier. A recent report[839] suggests that selected patients with a broad spectrum of congenital heart disease may enjoy same-day admission, limited sternotomy, immediate extubation, and very early discharge (25%, 74%, and 82% were discharged, respectively, at <24, <48, <72 hours from admission). It is unclear whether such efforts to reduce length of stay may have unintended negative consequences, such as increased complications and readmissions.

**Prior use.** Pediatric cardiac surgery mortality has not been widely used as an indicator of quality.

### *Empirical Evidence*

| Test | Statistic | Rating |
|---|---|---|
| *Precision* | | |
| Raw provider level rate/standard deviation | 7.2%, 1.7% | |
| Systematic provider -level standard deviation** | 1.5% | Moderate |
| Provider variation as a percentage of total variation** | 0.3% | Moderate |
| Signal ratio** | 22.2% | Low |
| R-Square** | 37.9% | Low |
| **APR-DRG, age-, gender- adjusted | | |
| *Minimum Bias – APR-DRG risk adjustment* | | |
| Signal variance change with risk adjustment | Decreases | Fair |
| Absolute impact: | | |
| Average absolute change (in %) | 12.8% | Good |
| Relative impact: | | |
| Rank correlation | 0.674 | Fair |
| Percent remaining in high decile/low decile | 16.7% / 66.7% | Fair |
| Percent changing more than 2 deciles | 35.1% | Fair |

**Precision.** This indicator is adequately precise, with a raw provider level mean of 7.2% and a substantial standard deviation of 1.7%. The systematic provider level standard deviation is moderate, at 1.5%. The provider level variation also accounts for a moderate percentage of total variation, at 0.3%. This means that relative to other indicators, a lower percentage of the variation occurs at the provider level, rather than the discharge level. The signal ratio is low, at 22.2%. This means that it is very likely that some the observed differences in provider performance do not represent true differences in provider performance. The R-square is substantially higher, but remains low relative to other indicators.

**Bias.** Signal variance decreases by more than 25% with risk adjustment, suggesting that some of the observed variance is due to differences in patient characteristics. The indicator performs fairly on the multiple measures of minimum bias. The rank correlation is fair at 0.674. The impact on the extremes is large, especially in the highest decile. Only 16.7% of providers in the highest decile remain, and only 66.7% in the lowest decile remain, after risk adjustment. Similarly, the number of providers moving at least two deciles in relative rank is also high. However, the absolute magnitude of risk adjustment is moderate.

**Construct validity.** Pediatric heart surgery does not load substantially on any of the three retained factors.

### *Discussion*

Pediatric heart surgeries include a diverse set of operations ranging from fairly straightforward to rather complex procedures. The mortality for the set of operations has been used as an outcome measure in the volume-outcome literature. Higher volume hospitals have been noted to have lower mortality rates, suggesting some difference in the processes of care between lower and higher volume hospitals, resulting in better outcomes. What those processes are, if they truly exist, is not known.

Relatively few hospitals perform pediatric heart surgery with frequency. This could effect the precision for many hospitals with low volumes. Our empirical tests found this indicator has lower precision relative to most other indicators. The signal ratio is low, indicating that some of the observed differences may not reflect true differences in performance. Multivariate techniques

do improve the ability to extract signal for this indicator somewhat, though the proportion that is extractable remains low relative to other indicators. Smoothing is recommended for this indicator to prevent the misidentification of outliers due to random noise.

Given the large variety in operations, and the varying risks associated with each, as well as the rather heterogeneous population receiving heart surgeries, it is likely that risk adjustment will be very important. Many volume-outcome studies used complicated clinical models for risk adjustment. Such adjustment is not available using APR-DRGs. Our empirical tests found substantial bias for this indicator, especially for providers with the highest mortality rates. Further, it is likely that given the complex mixture of procedures included in this definition, that APR-DRG risk adjustment is not adequate and that providers may need to supplement. Providers who wish to use this indicator may consider examining their case mixes, as well as the breakdown in the types of surgeries performed. Medical chart review may be helpful in determining whether more detailed risk adjustment affects hospital performance.

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

As with other mortality measures, there is concern that measuring mortality rates would result in access problems for higher risk patients. If possible, it may be useful to monitor whether operative risk declines with the implementation of this indicator.

Overall, this indicator is recommended for inclusion in the HCUP II QI set, when used in conjunction with volume measures. It received an empirical rating of 3 out of 26, and smoothing is highly recommended. This indicator is recommended with two major caveats of use. First, risk adjustment for clinical factors is recommended due to the substantial confounding bias for this indicator. Second, evidence supporting the construct validity of this indicator is limited.

# Chapter 4: Conclusions and Future Research

Our review of the literature and empirical performance of potential HCUP II quality indicators suggested that HCUP data can play a useful role in policy and management activities to improve quality, and might play an even more extensive role in the future. For quality indicators based on volume of inpatient procedures, HCUP is an excellent data source. For some other specific quality indicators related to the utilization of inpatient services, HCUP data can also provide some important insights. For example, because the "appropriate" rates of procedures like incidental appendectomy and bilateral cardiac catheterization are low, high hospital rates on these recommended HCUP II indicators are very likely to represent inappropriate care. For a range of other hospital quality indicators related to important outcomes - mortality for inpatient conditions and inpatient procedures – HCUP data reveal substantial differences across hospitals. If the high-rate hospitals could achieve the mortality rates of the lower-rate hospitals, literally thousands of inpatient deaths per year might be avoided. Our review also identified some potentially quite valuable uses of HCUP data for detecting important differences in the quality of outpatient care, as reflected in inpatient admissions for a range of chronic conditions that can usually be managed effectively on an ambulatory basis, and for acute complications that can be prevented through good ambulatory care.

Despite the promising findings of our review, we also identified some important limitations of indicators that could be constructed from HCUP data. Some of these limitations could be addressed through additional action and research. The limitations include:

(1) Some indicators, such as hospital volume indicators, are only proxies for quality. For example, while on average hospitals with higher volumes of certain complex procedures have been shown to have better outcomes, procedure volume is only weakly correlated with mortality and other outcomes, and volume indicators may encourage inappropriate utilization. It is important that the volume-outcome relationship or other relationships on which proxy measures are based be revisited to assure the validity of these indicators.

(2) A few indicators, such as laparoscopic cholecystectomy and congestive heart failure, may be susceptible to selection bias because the cases ascertainable from HCUP data do not represent the universe of patients with that condition or procedure. Injudicious use of these indicators may lead health care providers to admit patients who do not actually require inpatient care. A related problem is that inadequate or variable coding of key diagnoses may interfere with ascertainment of cases, such as for vaginal births after cesarean delivery.

(3) Missing information about patient outcomes is a potential source of bias for several mortality indicators, because 30-day mortality for some conditions substantially exceeds inpatient mortality. As a result, injudicious use of these indicators may lead health care providers to discharge patients prematurely, thereby shifting deaths to the outpatient setting.

(4) Patient characteristics, such as disease severity, comorbidities, physiologic derangements, and functional status, may substantially affect performance on mortality measures, and may vary systematically across providers or areas. In particular, confounders that cannot be identified from HCUP data, such as clinical information derived from physical examination, laboratory tests, and radiographic findings, and functional abnormalities, are of concern. Injudicious use of these indicators may lead health care providers to avoid high-risk patients, whose acuity is not fully captured by APR-DRGs or other severity classification systems, or to "upcode" comorbidities.

(5) Many indicators have limited evidence of construct validity, meaning that poor performance has not been clearly associated with poor quality of care (according to generally accepted process or risk-adjusted outcome indicators). Ideally, such associations should be demonstrated at both the patient and hospital or area levels. Without such evidence, providers or areas may find it difficult to address performance problems and identify opportunities for process improvement.

(6) Empirical analyses demonstrated that many indicators are somewhat imprecise, meaning that there is significant random variation making it difficult to discern substantial systematic effects at the provider level. The recommended indicators all demonstrated sufficient precision for use as quality indicators, but some benefit substantially from smoothing techniques. Using indicators without such techniques may lead to inappropriate conclusions about provider performance.

Given these limitations, quality indicators derived from currently available HCUP data are best used as "quality screens," meaning that they are particularly useful in identifying potential quality problems and identify areas for further investigation. These indicators may serve additional functions, as long as they are interpreted with consideration of the caveats noted in this report. For example, hospitals, hospital systems, hospital associations, local, state and Federal health agencies, managed care organizations, purchasers or consumer coalitions may use these indicators to compare health system performance with regional, national, or international benchmarks. Hospitals, hospital systems, or other provider representatives may use these indicators to reallocate resources among facilities or departments, or to develop and evaluate performance improvement activities. Regardless of the use of the QIs it is essential that the limitations of these measures be addressed.

Our principal findings suggest some directions for further research and development of HCUP and similar discharge-based data systems. To permit more confident conclusions based on HCUP II measures – that is, to improve the predictive power of HCUP II-based screens – additional research should be undertaken to assess the validity of promising measures. Many studies reviewed in this report have done this already, and have provided valuable evidence on the performance of discharge-based quality indicators. However, the previous studies have had a number of limitations. First, many studies focused on the validity of quality measures at the patient level, or whether patients who screen positive on the measure are more likely to have received poor quality care than patients who screen negative. But because these studies typically included only one or a few hospitals, they were unable to address a potentially more important question: at the level of *hospitals or areas*, does poor performance on an HCUP II indicator reflect a true quality problem? In other words, do hospitals with poorer performance on an HCUP II indicator provide lower quality of care than those with better performance on the same indicator? Such studies would require combining detailed process-of-care data from multiple hospitals with quality indicators based on administrative data (ideally with smoothing methods and risk adjustment) for the same group of hospitals and patient admissions. Although the cost of these investigations might seem prohibitive, there are efforts proposed, underway, and even completed to collect clinically-detailed data for a large and reasonably representative set of hospitals. For example, HCFA's Cooperative Cardiovascular Project (CCP) collected comprehensive chart review data for all heart attack patients in the traditional Medicare program admitted to virtually all hospitals in the United States during a 9-month period in 1994-95. Thus, at relatively low cost, it would be possible to extend recent work comparing hospital performance in heart attack care based on detailed CCP process measures to hospital

performance based on recommended HCUP II indicators for the same patients. Such studies would not only be useful for assessing the performance of HCUP II measures, but also for identifying where supplemental clinical data collection is most needed to evaluate quality of care.

Second, more research is needed to assess whether and how alternative risk adjustment methods affect indicator performance. Once again, many useful studies reviewed in our report have already been performed. Often, they found that a hospital's measured performance was not only sensitive to whether risk adjustment was performed, but also to which risk adjustment system was used. However, such studies have not been performed for most of the recommended indicators. Moreover, these studies usually did not take account of the improving data now available, including up to 25 fields for secondary diagnoses and flag variables (in New York and California) that distinguish between comorbidities present on admission (which generally should be included in risk adjustment) and complications that develop after admission (which generally should not). In addition, risk adjustment methods for administrative data are improving, which will benefit both quality measures and reimbursement systems. Finally, the previous studies generally did not account for random variation across hospitals in comparing different risk adjustment methods. Because such random variation contributes substantially to the apparent variation across hospitals for most HCUP II quality indicators, it is not surprising that relationships across risk adjustment systems are often weak. Simply put, the problem of serious "measurement error" complicates the detection of significant, consistent relationships. Future research using newer HCUP data, including the software accompanying this report that generates smoothed estimates, could address these limitations.

Even if these limitations of previous research are addressed in future studies, it seems likely that HCUP II indicators will generally remain imperfect measures of quality due to the constraints of the discharge data from which they are constructed. However, it is conceivable that many of these data limitations could be addressed in future versions of HCUP. One weakness involves the non-longitudinal nature of HCUP data; thus, a hospital may lower its inpatient mortality rate by transferring patients to other facilities or other services. Similarly, it is impossible to detect serious complications of treatment that occur after a patient is discharged. These weaknesses could be addressed by encouraging states to provide longitudinally linkable data to HCUP. In fact, many of the states participating in HCUP – including such states as California, Florida, Maryland, New York, and Washington – already collect unique patient identifiers or a combination of variables that can be used to create synthetic identifiers. A second weakness involves the lack of information on important health outcomes, particularly post-discharge mortality. Because of this weakness, it is impossible to track longer-term outcomes that are important for some conditions (e.g., hip fracture), nor is it possible to determine whether differences in discharge policies influence measured mortality rates. These weaknesses could be addressed by encouraging states to provide linked hospital discharge and death index data, or data with personal identifiers that could be linked to the National Death Index.

A third weakness involves the absence of outpatient data; as all but the most intensive therapies continue to move out of the inpatient setting, inpatient data alone provide an increasingly limited view of overall quality of care. With data on ambulatory surgery activity, a much broader range of procedure-related quality indicators would be feasible in HCUP. These would include volume, utilization, and complication or readmission indicators for cardiac catheterization, electrophysiologic stimulation, and a range of arthroscopic and gynecologic

procedures. In addition, more complete indicators could be developed for such procedures as angioplasty and laparoscopic cholecystectomy. Nine HCUP states are already collecting such data. With data on emergency department activity, more complete and accurate indicators for quality of care for serious conditions that are sometimes managed on an outpatient basis could be constructed. These conditions include pneumonia, congestive heart failure, angina, and stroke; in the future, acute myocardial infarction might also be included. Emergency department data would also support the construction of more complete quality indicators related to ambulatory care sensitive conditions. One HCUP state is piloting such data collection; expanding this activity to other states is an important goal. In addition to outpatient hospital data, physician office data may provide important information regarding care previous to hospital admissions and other quality. This data could provide a more complete picture of quality for many indicators, particularly the ACSC indicators, where outpatient office care is critical. Nonetheless, to our knowledge the effort to link hospitalizations with outpatient physician office data is far from fruition.

A longer-term but potentially realistic goal is to incorporate more clinical information into HCUP. One approach to doing this would involve data abstracted from electronic medical records. Because such records are neither widespread nor standardized yet, an appropriate focus of research might be to facilitate the adoption of electronic records, and especially to encourage the development of standardized terminology and data structures. It would be helpful to identify a key set of clinical data elements that should be readily abstractable from any electronic medical record system. Other sources of clinical data may be more promising in the shorter run. These include disease registries, such as national registries (e.g. Surveillance, Epidemiology, and End Results (SEER) registries) and state registries of cancer cases. Such data have already been linked to Medicare records, and have proven extremely valuable in providing insights into the quality of care received by cancer patients. For example, the studies cited in this report on volume-outcome relationships for complex cancer surgeries were based in large part on linked SEER-Medicare data. Another potential source is electronic laboratory information, which for some indicators permits risk adjustment that performs about as well as detailed chart review data. While additional clinical detail in HCUP is certainly a longer-term prospect, it seems clear that electronic clinical data will become a much more central feature of health care delivery in the future. It is not too early to support research to explore the consequences of this trend for HCUP and other electronic data systems that have traditionally been viewed as "administrative."

Many of these recommendations for further HCUP research and development involve enriching HCUP data, or linking it for research purposes to clinically detailed datasets. But such enrichment also leads to increased concerns about data confidentiality and security. Misuse of such data, including the inadvertent or deliberate identification of individual medical histories, is a risk that should not be underestimated. Thus, in contrast to a largely "public use" policy for current HCUP data, such enriched data would require stricter procedures to assure confidentiality. Models for how such data could be handled are provided by HCFA for Medicare data and by some of the HCUP states for the "enriched" versions of their datasets. Briefly, the use of more detailed data requires a progressively more careful review process, to assure that the benefits of the proposed research sufficiently outweigh any associated risks , and to assure that the researchers adhere to strict data security protocols (e.g., secure servers, steps to remove confidential information as early in the research process as feasible, IRB evaluation, etc.). HCUP is in a somewhat difficult position on this topic, as its dependence on data produced by states means that data use and security protocols must meet state standards. But AHRQ is in an

excellent position to work with states to help them develop appropriate protocols if they have not done so already, and to develop a set of standards for using data from multiple states that meets all state requirements. This is a critical data use issue that goes beyond HCUP: to gain a more complete picture of quality of care, especially for non-Medicare patients, it will be necessary to use data related to quality from multiple states and/or payers. If possible, federal financial support for such activities would be appropriate. The benefits of analyzing data from particular states or payers do not accrue only to their own patients, because such data provide a more complete picture of variation in quality of care that can help to establish benchmarks and guidance for the rest of the country. Moreover, this is a critical time for leadership on the issue of use of confidential data for research purposes. On the one hand, the information involved is very sensitive; on the other hand, to the extent that data from the whole population are not included in studies of health care quality, it is impossible to reach reliable conclusions about opportunities for improving quality of care. The HCUP project provides an opportunity to demonstrate effective ways to balance these concerns, and could be a model for other such collaborative efforts to improve data and research on quality of care and other critical issues affecting the nation's health.

# References


**1. Institute of Medicine Division of Health Care Services. Medicare: a strategy for quality assurance. Washington, D.C.: National Academy Press; 1990.**

**2. Chassin MR, Galvin RW. The urgent need to improve health care quality. Institute of Medicine National Roundtable on Health Care Quality. JAMA 1998;280(11):1000-5.**

**3. Johantgen M, Elixhauser A, Bali JK, et al. Quality indicators using hospital discharge data: state and national applications. Jt Comm J Qual Improv 1998;24(2):88-105.**

**4. Version 1 HCUP Quality Indicators. 1 ed: Agency for Health Care Policy and Research; 1994.**

**5. Healthy People 2010. ,Office of Disease Prevention and Health Promotion U.S. Department of Health and Human Services.**

**6. McGlynn EA, Asch SM. Developing a clinical performance measure. Am J Prev Med 1998;14(3 Suppl):14-21.**

**7. Iezzoni LI, Ash AS, Coffman GA, et al. Predicting in-hospital mortality. A comparison of severity measurement approaches. Med Care 1992;30(4):347-59.**

**8. Iezzoni LI, editor. Risk Adjustment for Measuring Health Outcomes. Chicago: Health Administration Press; 1994.**

**9. Iezzoni LI, Ash AS, Shwartz M, et al. Predicting who dies depends on how severity is measured: implications for evaluating patient outcomes. Ann Intern Med 1995;123(10):763-70.**

**10. Iezzoni LI, Shwartz M, Ash AS, et al. Predicting in-hospital mortality for stroke patients: results differ across severity-measurement methods. Med Decis Making 1996;16(4):348-56.**

**11. Iezzoni LI, Shwartz M, Ash AS, et al. Using severity-adjusted stroke mortality rates to judge hospitals. Int J Qual Health Care 1995;7(2):81-94.**

**12. Iezzoni LI, Ash AS, Shwartz M, et al. Judging hospitals by severity-adjusted mortality rates: the influence of the severity-adjustment method. Am J Public Health 1996;86(10):1379-87.**

**13. Iezzoni LI, Shwartz M, Ash AS, et al. Severity measurement methods and judging hospital death rates for pneumonia. Med Care 1996;34(1):11-28.**

**14. Iezzoni LI. Assessing quality using administrative data. Ann Intern Med 1997;127(8 Pt 2):666-74.**

**15. Iezzoni LI. The risks of risk adjustment. JAMA 1997;278(19):1600-7.**

**16. Iezzoni LI. Risk adjustment for measuring healthcare outcomes. 2nd ed. Chicago, Ill.: Health Administration Press; 1997.**

**17. Iezzoni LI, Ash AS, Shwartz M, et al. Predicting in-hospital deaths from coronary artery bypass graft surgery. Do different severity measures give different predictions? Med Care 1998;36(1):28-39.**

**18. Iezzoni LI, Shwartz M, Ash AS, et al. Using severity measures to predict the likelihood of death for pneumonia inpatients. J Gen Intern Med 1996;11(1):23-31.**

**19. Pine M, Norusis M, Jones B, et al. Predictions of hospital mortality rates: a comparison of data sources. Ann Intern Med 1997;126(5):347-54.**

**20. Siu AL, McGlynn EA, Morgenstern H, et al. Choosing quality of care measures based on the expected impact of improved care on health. Health Serv Res 1992;27(5):619-50.**
**21. Donabedian A. Explorations in Quality Assessment and Monitoring.The definition of quality and approaches to its assessment. Ann Arbor, MI: Health Administration Press; 1980.**
**22. Donabedian A. The quality of care: how can it be assessed? JAMA 1988;260(12):1743-1748.**
**23. Schneider EC, Epstein AM. Influence of cardiac-surgery performance reports on referral practices and access to care. A survey of cardiovascular specialists. N Engl J Med 1996;335(4):251-6.**
**24. Mennemeyer ST, Morrisey MA, Howard LZ. Death and reputation: how consumers acted upon HCFA mortality information. Inquiry 1997;34(2):117-28.**
**25. Hibbard JH, Jewett JJ. Will quality report cards help consumers? Health Aff (Millwood) 1997;16(3):218-28.**
**26. Normand SL, McNeil BJ, Peterson LE, et al. Eliciting expert opinion using the Delphi technique: identifying performance indicators for cardiovascular disease. Int J Qual Health Care 1998;10(3):247-60.**
**27. Veroff DR, Gallagher PM, Wilson V, et al. Effective reports for health care quality data: lessons from a CAHPS demonstration in Washington State. Int J Qual Health Care 1998;10(6):555-60.**
**28. Delbanco TL, Stokes DM, Cleary PD, et al. Medical patients' assessments of their care during hospitalization: insights for internists. J Gen Intern Med 1995;10(12):679-85.**
**29. Laine C, Davidoff F. Patient-centered medicine. A professional evolution. JAMA 1996;275(2):152-6.**
**30. Allen HM, Jr., Rogers WH. Consumer surveys of health plan performance: a comparison of content and approach and a look to the future. Jt Comm J Qual Improv 1996;22(12):775-94.**
**31. Cleary PD, Edgman-Levitan S. Health care quality. Incorporating consumer perspectives. JAMA 1997;278(19):1608-12.**
**32. Eye on patients: excerpts from a report on patients' concerns and experiences about the health care system. American Hospital Association and the Picker Institute. J Health Care Finance 1997;23(4):2-11.**
**33. Calnan MW. The patient's perspective. Int J Technol Assess Health Care 1998;14(1):24-34.**
**34. Tye L. Patient surveys show how Massachusetts hospitals stack up. Boston Globe 1998 Novemebr 13, 1998;Sect. A1, A34.**
**35. Finkelstein BS, Harper DL, Rosenthal GE. Patient assessments of hospital maternity care: a useful tool for consumers? Health Serv Res 1999;34(2):623-40.**
**36. Berwick DM, Wald DL. Hospital leaders' opinions of the HCFA mortality data. JAMA 1990;263(2):247-9.**
**37. Jencks SF, Daley J, Draper D, et al. Interpreting hospital mortality data. The role of clinical risk adjustment. JAMA 1988;260(24):3611-6.**
**38. Park RE, Brook RH, Kosecoff J, et al. Explaining variations in hospital death rates. Randomness, severity of illness, quality of care. JAMA 1990;264(4):484-90.**

**39. Localio AR, Hamory BH, Sharp TJ, et al. Comparing hospital mortality in adult patients with pneumonia. A case study of statistical methods in a managed care program. Ann Intern Med 1995;122(2):125-32.**
**40. Localio AR, Hamory BH, Fisher AC, et al. The public release of hospital and physician mortality data in Pennsylvania. A case study. Med Care 1997;35(3):272-286.**
**41. Hofer TP, Hayward RA. Identifying poor-quality hospitals. Can hospital mortality rates detect quality problems for medical diagnoses? Med Care 1996;34(8):737-53.**
**42. Thomas JW, Hofer TP. Accuracy of risk-adjusted mortality rate as a measure of hospital quality of care. Med Care 1999;37(1):83-92.**
**43. Mant J, Hicks N. Detecting differences in quality of care: the sensitivity of measures of process and outcome in treating acute myocardial infarction. BMJ 1995;311(7008):793-6.**
**44. Palmer RH. Process-based measures of quality: the need for detailed clinical data in large health care databases. Ann Intern Med 1997;127(8 Pt 2):733-8.**
**45. Eddy DM. Performance measurement: problems and solutions . Health Aff (Millwood) 1998;17(4):7-25.**
**46. Harr DS, Balas EA, Mitchell J. Developing quality indicators as educational tools to measure the implementation of clinical practice guidelines. Am J Med Qual 1996;11(4):179-85.**
**47. Ellerbeck EF, Jencks SF, Radford MJ, et al. Quality of care for Medicare patients with acute myocardial infarction. A four-state pilot study from the Cooperative Cardiovascular Project . JAMA 1995;273(19):1509-14.**
**48. Marciniak TA, Ellerbeck EF, Radford MJ, et al. Improving the quality of care for Medicare patients with acute myocardial infarction: results from the Cooperative Cardiovascular Project . JAMA 1998;279(17):1351-7.**
**49. Donabedian A. Evaluating the quality of medical care. Milbank Mem Fund Q 1966;44(3):Suppl:166-206.**
**50. Brook RH, McGlynn EA, Cleary PD. Quality of health care. Part 2: measuring quality of care. N Engl J Med 1996;335(13):966-70.**
**51. Luft HS, Bunker JP, Enthoven AC. Should operations be regionalized? The empirical relation between surgical volume and mortality. N Engl J Med 1979;301(25):1364-9.**
**52. Hughes RG, Garnick DW, Luft HS, et al. Hospital volume and patient outcomes. The case of hip fracture patients. Med Care 1988;26(11):1057-67.**
**53. Hannan EL, JF OD, Kilburn H, Jr., et al. Investigation of the relationship between volume and mortality for surgical procedures performed in New York State hospitals. JAMA 1989;262(4):503-10.**
**54. Hannan EL, Kilburn H, Jr., Bernard H, et al. Coronary artery bypass surgery: the relationship between inhospital mortality rate and surgical volume after controlling for clinical risk factors. Med Care 1991;29(11):1094-107.**
**55. Stone VE, Seage GRd, Hertz T, et al. The relation between hospital experience and mortality for patients with AIDS. JAMA 1992;268(19):2655-61.**
**56. Hosenpud JD, Breen TJ, Edwards EB, et al. The effect of transplant center volume on cardiac transplant outcome. A report of the United Network for Organ Sharing Scientific Registry. JAMA 1994;271(23):1844-9.**

**57. Jones A, O'Driscoll K, Luke LC. Head injuries and the observation ward [letter; comment]. J Accid Emerg Med 1995;12(2):160-1.**
**58. Phibbs CS, Bronstein JM, Buxton E, et al. The effects of patient volume and level of care at the hospital of birth on neonatal mortality. JAMA 1996;276(13):1054-9.**
**59. Ellis SG, Weintraub W, Holmes D, et al. Relation of operator volume and experience to procedural outcome of percutaneous coronary revascularization at hospitals with high interventional volumes. Circulation 1997;95(11):2479-84.**
**60. Jollis JG, Peterson ED, Nelson CL, et al. Relationship between physician and hospital coronary angioplasty volume and outcome in elderly patients. Circulation 1997;95(11):2485-91.**
**61. Hannan EL, Racz M, Ryan TJ, et al. Coronary angioplasty volume-outcome relationships for hospitals and cardiologists. JAMA 1997;277(11):892-8.**
**62. Dardik A, Burleyson GP, Bowman H, et al. Surgical repair of ruptured abdominal aortic aneurysms in the state of Maryland: factors influencing outcome among 527 recent cases. J Vasc Surg 1998;28(3):413-20.**
**63. Rosenthal GE, Shah A, Way LE, et al. Variations in standardized hospital mortality rates for six common medical diagnoses: implications for profiling hospital quality. Med Care 1998;36(7):955-64.**
**64. Cebul RD, Snow RJ, Pine R, et al. Indications, outcomes, and provider volumes for carotid endarterectomy. JAMA 1998;279(16):1282-7.**
**65. Begg CB, Cramer LD, Hoskins WJ, et al. Impact of hospital volume on operative mortality for major cancer surgery. JAMA 1998;280(20):1747-51.**
**66. Thiemann DR, Coresh J, Oetgen WJ, et al. The association between hospital volume and survival after acute myocardial infarction in elderly patients . N Engl J Med 1999;340(21):1640-8.**
**67. Pratt R, Burr G, Leelarthaepin B, et al. The effects of All-RN and RN-EN staffing on the quality and cost of patient care. Aust J Adv Nurs 1993;10(3):27-39.**
**68. Archibald LK, Manning ML, Bell LM, et al. Patient density, nurse-to-patient ratio and nosocomial infection risk in a pediatric cardiac intensive care unit. Pediatr Infect Dis J 1997;16(11):1045-8.**
**69. Blegen MA, Goode CJ, Reed L. Nurse staffing and patient outcomes. Nurs Res 1998;47(1):43-50.**
**70. Czaplinski C, Diers D. The effect of staff nursing on length of stay and mortality. Med Care 1998;36(12):1626-38.**
**71. Kovner C, Gergen PJ. Nurse staffing levels and adverse events following surgery in U.S. hospitals. Image J Nurs Sch 1998;30(4):315-21.**
**72. McCloskey JM. Nurse staffing and patient outcomes. Nurs Outlook 1998;46(5):199-200.**
**73. Bond CA, Raehl CL, Pitterle ME, et al. Health care professional staffing, hospital characteristics, and hospital mortality rates. Pharmacotherapy 1999;19(2):130-8.**
**74. Wennberg J, Gittelsohn. Small area variations in health care delivery. Science 1973;182(117):1102-8.**
**75. Wennberg J, Gittelsohn A. Variations in medical care among small areas. Sci Am 1982;246(4):120-34.**

**76. Markowitz JS, Pashko S, Gutterman EM, et al. Death rates among patients hospitalized with community-acquired pneumonia: a reexamination with data from three states. Am J Public Health 1996;86(8 Pt 1):1152-4.**
**77. Hofer TP, Wolfe RA, Tedeschi PJ, et al. Use of community versus individual socioeconomic data in predicting variation in hospital use. Health Serv Res 1998;33(2 Pt 1):243-59.**
**78. Jencks SF, Dobson A. Refining case-mix adjustment. The research evidence. N Engl J Med 1987;317(11):679-86.**
**79. Shapiro MF, Park RE, Keesey J, et al. The effect of alternative case-mix adjustments on mortality differences between municipal and voluntary hospitals in New York City. Health Serv Res 1994;29(1):95-112.**
**80. Halm EA, Lee C, Chassin MR. How is volume related to quality in health care? A systematic review of the research literature: Institute of Medicine, National Academy of Sciences, Division of Health Care Services, Committee on Quality of Care in America; 2000 May 1.**
**81. Thomas J, Holloway J, Guire K. Validating risk-adjusted mortality as an indicator for quality of care. Inquiry 1993;30(1):6-22.**
**82. Hofer TP, Hayward RA. Can early re-admission rates accurately detect poor-quality hospitals? Med Care 1995;33(3):234-45.**
**83. Thomas JW. Does risk-adjusted readmission rate provide valid information on hospital quality? Inquiry 1996;33(3):258-70.**
**84. Normand ST, Glickman ME, Sharma RG, et al. Using admission characteristics to predict short-term mortality from myocardial infarction in elderly patients. Results from the Cooperative Cardiovascular Project . JAMA 1996;275(17):1322-8.**
**85. Thomas JW, Hofer TP. Research evidence on the validity of risk-adjusted mortality rate as a measure of hospital quality of care. Med Care Res Rev 1998;55(4):371-404.**
**86. 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;281(22):2098-105.**
**87. Normand S, Glickman M, Gastonis C. Statistical methods for profiling providers of medical care: Issues and applications. JASA 1997;92(439):803-814.**
**88. O'Hagan A. Bayesian Inference. In: al. GPe, editor. Kendall's Advanced Theory of Statistics. New York: Halstead Press; 1994.**
**89. Goldstein H. Multilevel Statistical Models. 2nd ed. New York: Halstead Press; 1995.**
**90. Wennberg JE, Freeman JL, Shelton RM, et al. Hospital use and mortality among Medicare beneficiaries in Boston and New Haven. N Engl J Med 1989;321(17):1168-73.**
**91. Fisher ES, Wennberg JE, Stukel TA, et al. Hospital readmission rates for cohorts of Medicare beneficiaries in Boston and New Haven. N Engl J Med 1994;331(15):989-95.**
**92. Miller MG, Miller LS, Fireman B, et al. Variation in practice for discretionary admissions. Impact on estimates of quality of hospital care . JAMA 1994;271(19):1493-8.**
**93. Rosenthal GE, Harper DL, Shah A, et al. A regional evaluation of variation in low-severity hospital admissions. J Gen Intern Med 1997;12(7):416-22.**
**94. Fisher ES, Wennberg JE, Stukel TA, et al. Associations among hospital capacity, utilization, and mortality of US Medicare beneficiaries, controlling for sociodemographic factors. Health Serv Res 2000;34(6):1351-62.**

**95. McClellan M, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. JAMA 1994;272(11):859-66.**
**96. Iezzoni LI, Heeren T, Foley SM, et al. Chronic conditions and risk of in-hospital death. Health Serv Res 1994;29(4):435-60.**
**97. Jones RH, Hannan EL, Hammermeister KE, et al. Identification of preoperative variables needed for risk adjustment of short-term mortality after coronary artery bypass graft surgery. The Working Group Panel on the Cooperative CABG Database Project. J Am Coll Cardiol 1996;28(6):1478-87.**
**98. Khuri SF, Daley J, Henderson W, et al. Risk adjustment of the postoperative mortality rate for the comparative assessment of the quality of surgical care: results of the National Veterans Affairs Surgical Risk Study. J Am Coll Surg 1997;185(4):315-27.**
**99. Wray NP, Hollingsworth JC, Peterson NJ, et al. Case-mix adjustment using administrative databases: a paradigm to guide future research. Med Care Res Rev 1997;54(3):326-56.**
**100. Kuttner R. The risk-adjustment debate. N Engl J Med 1998;339(26):1952-6.**
**101. Block PC, Peterson EC, Krone R, et al. Identification of variables needed to risk adjust outcomes of coronary interventions: evidence-based guidelines for efficient data collection. J Am Coll Cardiol 1998;32(1):275-82.**
**102. Richardson D, Tarnow-Mordi WO, Lee SK. Risk adjustment for quality improvement. Pediatrics 1999;103(1 Suppl E):255-65.**
**103. Iezzoni LI, Foley SM, Daley J, et al. Comorbidities, complications, and coding bias. Does the number of diagnosis codes matter in predicting in-hospital mortality? JAMA 1992;267(16):2197-203.**
**104. Green J, Wintfeld N. How accurate are hospital discharge data for evaluating effectiveness of care? Med Care 1993;31(8):719-31.**
**105. Malenka DJ, McLerran D, Roos N, et al. Using administrative data to describe casemix: a comparison with the medical record. J Clin Epidemiol 1994;47(9):1027-32.**
**106. Jencks SF, Williams DK, Kay TL. Assessing hospital-associated deaths from discharge data. The role of length of stay and comorbidities. JAMA 1988;260(15):2240-6.**
**107. Romano PS, Mark DH. Bias in the coding of hospital discharge data and its implications for quality assessment. Med Care 1994;32(1):81-90.**
**108. Simborg DW. DRG creep: a new hospital-acquired disease. N Engl J Med 1981;304(26):1602-4.**
**109. Keeler EB, Kahn KL, Draper D, et al. Changes in sickness at admission following the introduction of the prospective payment system. JAMA 1990;264(15):1962-8.**
**110. Hsia DC, Krushat WM, Fagan AB, et al. Accuracy of diagnostic coding for Medicare patients under the prospective-payment system. N Engl J Med 1988;318(6):352-355.**
**111. Green J, Wintfeld N. Report cards on cardiac surgeons. Assessing New York State's approach. N Engl J Med 1995;332(18):1229-32.**
**112. Hannan EL, Racz MJ, Jollis JG, et al. Using Medicare claims data to assess provider quality for CABG surgery: does it work well enough? Health Serv Res 1997;31(6):659-78.**
**113. Romano PS, Chan BK. Risk-adjusting acute myocardial infarction mortality: are APR-DRGs the right tool? Health Serv Res 2000;34(7):1469-89.**

**114. Goldfield N, Averill R. On "risk-adjusting acute myocardial infarction mortality: are APR-DRGs the right tool"? [comment]. Health Serv Res 2000;34(7):1491-5; discussion 1495-8.**
**115. Jollis JG, Romano PS. Pennsylvania's Focus on Heart Attack--grading the scorecard. N Engl J Med 1998;338(14):983-7.**
**116. O'Connor GT, Plume SK, Olmstead EM, et al. Multivariate prediction of in-hospital mortality associated with coronary artery bypass graft surgery. Northern New England Cardiovascular Disease Study Group. Circulation 1992;85(6):2110-8.**
**117. O'Connor GT, Plume SK, Olmstead EM, et al. A regional intervention to improve the hospital mortality associated with coronary artery bypass graft surgery. The Northern New England Cardiovascular Disease Study Group. JAMA 1996;275(11):841-6.**
**118. Hannan EL, Kilburn H, Jr., Racz M, et al. Improving the outcomes of coronary artery bypass surgery in New York State. JAMA 1994;271(10):761-6.**
**119. Knaus WA, Draper EA, Wagner DP, et al. An evaluation of outcome from intensive care in major medical centers. Ann Intern Med 1986;104(3):410-8.**
**120. Knaus WA, Wagner DP, Zimmerman JE, et al. Variations in mortality and length of stay in intensive care units. Ann Intern Med 1993;118(10):753-61.**
**121. Fine MJ, Auble TE, Yealy DM, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. N Engl J Med 1997;336(4):243-50.**
**122. Cooper GS, Chak A, Harper DL, et al. Care of patients with upper gastrointestinal hemorrhage in academic medical centers: a community-based comparison. Gastroenterology 1996;111(2):385-90.**
**123. Krumholz HM, Chen J, Wang Y, et al. Comparing AMI mortality among hospitals in patients 65 years of age and older: evaluating methods of risk adjustment. Circulation 1999;99(23):2986-92.**
**124. Luft H, Romano P, Remy L, et al. Second Report of the California Hospital Outcomes Project: Acute Myocardial Infarction. ,Office of Statewide Health Planning and Development.**
**125. Iezzoni LI, Ash AS, Shwartz M, et al. Differences in procedure use, in-hospital mortality, and illness severity by gender for acute myocardial infarction patients: are answers affected by data source and severity measure? Med Care 1997;35(2):158-71.**
**126. Hibbard JH, Jewett JJ, Legnini MW, et al. Choosing a health plan: do large employers use the data? Health Aff (Millwood) 1997;16(6):172-80.**
**127. Hibbard JH, Jewett JJ, Engelmann S, et al. Can Medicare beneficiaries make informed choices? Health Aff (Millwood) 1998;17(6):181-93.**
**128. Schneider EC, Epstein AM. Use of public performance reports: a survey of patients undergoing cardiac surgery. JAMA 1998;279(20):1638-42.**
**129. Booske BC, Sainfort F, Hundt AS. Eliciting consumer preferences for health plans. Health Serv Res 1999;34(4):839-54.**
**130. Rainwater JA, Romano PS, Antonius DM. The California Hospital Outcomes Project: how useful is California's report card for quality improvement? . Jt Comm J Qual Improv 1998;24(1):31-9.**
**131. Romano PS, Rainwater JA, Antonius D. Grading the graders: how hospitals in California and New York perceive and interpret their report cards. Med Care 1999;37(3):295-305.**

**132. Palmer RH, Louis TA, Peterson HF, et al. What makes quality assurance effective? Results from a randomized, controlled trial in 16 primary care group practices. Med Care 1996;34(9 Suppl):SS29-39.**
**133. Duffy SQ, Farley DE. Patterns of decline among inpatient procedures. Public Health Rep 1995;110(6):674-81.**
**134. Rutledge R. Can medical school-affiliated hospitals compete with private hospitals in the age of managed care? An 11-state, population-based analysis of 351,201 patients undergoing cholecystectomy. J Am Coll Surg 1997;185(3):207-17.**
**135. Maynard C, Chapko MK, Every NR, et al. Coronary angioplasty outcomes in the Healthcare Cost and Utilization Project, 1993-1994. Am J Cardiol 1998;81(7):848-52.**
**136. Shepardson LB, Youngner SJ, Speroff T, et al. Increased risk of death in patients with do-not-resuscitate orders. Med Care 1999;37(8):727-37.**
**137. Layde PM, Broste SK, Desbiens N, et al. Generalizability of clinical studies conducted at tertiary care medical centers: a population-based analysis. J Clin Epidemiol 1996;49(8):835-41.**
**138. Hannan EL, Kilburn H, Jr., O'Donnell JF, et al. Interracial access to selected cardiac procedures for patients hospitalized with coronary artery disease in New York State. Med Care 1991;29(5):430-41.**
**139. Buckle JM, Horn SD, Oates VM, et al. Severity of illness and resource use differences among white and black hospitalized elderly. Arch Intern Med 1992;152(8):1596-603.**
**140. McBean AM, Gornick M. Differences by race in the rates of procedures performed in hospitals for Medicare beneficiaries. Health Care Financ Rev 1994;15(4):77-90.**
**141. McBean AM, Warren JL, Babish JD. Continuing differences in the rates of percutaneous transluminal coronary angioplasty and coronary artery bypass graft surgery between elderly black and white Medicare beneficiaries. Am Heart J 1994;127(2):287-95.**
**142. Williams JF, Zimmerman JE, Wagner DP, et al. African-American and white patients admitted to the intensive care unit: is there a difference in therapy and outcome? Crit Care Med 1995;23(4):626-36.**
**143. Phillips RS, Hamel MB, Teno JM, et al. Race, resource use, and survival in seriously ill hospitalized adults. The SUPPORT Investigators. J Gen Intern Med 1996;11(7):387-96.**
**144. Romano PS, Campa DR, Rainwater JA. Elective cervical discectomy in California: postoperative in-hospital complications and their risk factors. Spine 1997;22(22):2677-92.**
**145. Huber TS, Wang JG, Wheeler KG, et al. Impact of race on the treatment for peripheral arterial occlusive disease. J Vasc Surg 1999;30(3):417-25.**
**146. Chen J, Radford MJ, Wang Y, et al. Do "America's Best Hospitals" perform better for acute myocardial infarction? N Engl J Med 1999;340:286-92.**
**147. Hartz AJ, Gottlieb MS, Kuhn EM, et al. The relationship between adjusted hospital mortality and the results of peer review. Health Serv Res 1993;27(6):765-77.**
**148. Hsia DC, Ahern CA, Ritchie BP, et al. Medicare reimbursement accuracy under the prospective**
**payment system, 1985 to 1988. JAMA 1992;268(7):896-899.**
**149. Cullen DJ, Bates DW, Small SD, et al. The incident reporting system does not detect adverse drug events: a problem for quality improvement. Jt Comm J Qual Improv 1995;21(10):541-8.**

**150. Kohn L, Corrigan J, Donaldson M, et al., editors. To Err Is Human: Building a Safer Health System. Washington, D.C.: National Academy Press; 2000.**
**151. Silber JH, Rosenbaum PR, Schwartz JS, et al. Evaluation of the complication rate as a measure of quality of care in coronary artery bypass graft surgery. JAMA 1995;274(4):317-23.**
**152. Silber JH, Rosenbaum PR, Ross RN. Comparing the contribution of predictors: which outcomes vary with hospital rather than patient characteristics? J Am Stat Assoc 1995;90:7-18.**
**153. Silber JH, Rosenbaum PR, Williams SV, et al. The relationship between choice of outcome measure and hospital rank in general surgical procedures: implications for quality assessment. Int J Qual Health Care 1997;9:193-200.**
**154. Kahn KL, Brook RH, Draper D, et al. Interpreting hospital mortality data. How can we proceed? JAMA 1988;260(24):3625-8.**
**155. Mullins RJ, Mann NC, Hedges JR, et al. Adequacy of hospital discharge status as a measure of outcome among injured patients. JAMA 1998;279(21):1727-31.**
**156. Sands K, Vineyard G, Platt R. Surgical site infections occurring after hospital discharge. J Infect Dis 1996;173(4):963-70.**
**157. Sands K, Vineyard G, Livingston J, et al. Efficient identification of postdischarge surgical site infections: use of automated pharmacy dispensing information, administrative data, and medical record information. J Infect Dis 1999;179(2):434-41.**
**158. Iezzoni LI, Mackiernan YD, Cahalane MJ, et al. Screening inpatient quality using post-discharge events. Med Care 1999;37(4):384-98.**
**159. Omoigui NA, Miller DP, Brown KJ, et al. Outmigration for coronary bypass surgery in an era of public dissemination of clinical outcomes. Circulation 1996;93(1):27-33.**
**160. Hannan EL, Siu AL, Kumar D, et al. Assessment of coronary artery bypass graft surgery performance in New York. Is there a bias against taking high-risk patients? Med Care 1997;35(1):49-56.**
**161. Petersen LA, Orav EJ, Teich JM, et al. Using a computerized sign-out program to improve continuity of inpatient care and prevent adverse events. Jt Comm J Qual Improv 1998;24(2):77-87.**
**162. Dranove. Information is good except when its bad. ,Stanford University Working Paper.**
**163. Marshall MN, Shekelle PG, Leatherman S, et al. The public release of performance data: what do we expect to gain? A review of the evidence. Jama 2000;283(14):1866-74.**
**164. Olfson M, Marcus S, Sackeim HA, et al. Use of ECT for the inpatient treatment of recurrent major depression. Am J Psychiatry 1998;155(1):22-9.**
**165. Meurer JR, Kuhn EM, George V, et al. Charges for childhood asthma by hospital characteristics. Pediatrics 1998;102(6):E70.**
**166. Lanska DJ, Hartz AJ. Measurement of quality in health care . Neurology 1998;50(3):584-7.**
**167. Lanska DJ, Kryscio RJ. In-hospital mortality following carotid endarterectomy. Neurology 1998;51(2):440-7.**
**168. Schnitzler MA, Lambert DL, Mundy LM, et al. Variations in healthcare measures by insurance status for patients receiving ventilator support. Clin Perform Qual Health Care 1998;6(1):17-22.**

**169. Niederman MS, McCombs JS, Unger AN, et al. The cost of treating community-acquired pneumonia. Clin Ther 1998;20(4):820-37.**
**170. Zhao SZ, Wong JM, Davis MB, et al. The cost of inpatient endometriosis treatment: an analysis based on the Healthcare Cost and Utilization Project Nationwide Inpatient Sample. Am J Manag Care 1998;4(8):1127-34.**
**171. Rentz AM, Halpern MT, Bowden R. The impact of candidemia on length of hospital stay, outcome, and overall cost of illness. Clin Infect Dis 1998;27(4):781-8.**
**172. Ritchie JL, Maynard C, Chapko MK, et al. Association between percutaneous transluminal coronary angioplasty volumes and outcomes in the Healthcare Cost and Utilization Project 1993-1994. Am J Cardiol 1999;83(4):493-7.**
**173. Best AE. Secondary data bases and their use in outcomes research: a review of the area resource file and the Healthcare Cost and Utilization Project. J Med Syst 1999;23(3):175-81.**
**174. Krumholz HM, Chen YT, Bradford WD, et al. Variations in and correlates of length of stay in academic hospitals among patients with heart failure resulting from systolic dysfunction. Am J Manag Care 1999;5(6):715-23.**
**175. Seifeldin R, Hantsch JJ. The economic burden associated with colon cancer in the United States. Clin Ther 1999;21(8):1370-9.**
**176. Williams GR, Jiang JG, Matchar DB, et al. Incidence and occurrence of total (first-ever and recurrent) stroke. Stroke 1999;30(12):2523-8.**
**177. Kirsner RS, Yang DG, Kerdel FA. Dermatologic disease accounts for a large number of hospital admissions annually. J Am Acad Dermatol 1999;41(6):970-3.**
**178. MEDLINE [database online]. Bethesda (MD),National Library of Medicine.1985-Updated weekly.Available from: National Library of Medicine; BRS Information Technologies, McLean, VA; DIALOG information Services, Inc., Palo Alto, CA.**
**179. PsycINFO. Washington (DC),American Psychological Association.1887-Updated monthly (journals), 1987-current (books and chapters).Available from: American Psycological Association, Washinton DC; Ovid Technologies, New York, NY; Silverplatter Information, Norwood, MA.**
**180. The Cochrane Library. Oxford, England,Update Software, Ltd.Available from: Update Software, San Diego, CA.**
**181. EMBASE. The Netherlands,Elsevier Science Publishers B.V.1974-Updated weekly.Available from: Elsevier Science, Secondary Publishing Division, New York, NY.**
**182. Ingber MJ. The current state of risk adjustment technology for capitation. J Ambulatory Care Manage 1998;21(4):1-28.**
**183. Rosenbaum P, Rubin D. Reducing bias in observational studies using subclassification on the propensity score. J Am Stat Assoc 1894;79:516-24.**
**184. Glazer J, McGuire T. Minimum variance optimal risk adjustment. American Economic Review 1999;In Press.**
**185. Chamberlain G. Panel Data. In: Griliches Z, Intriligator M, editors. Handbook of Economics. New York: Elsevier Science; 1984. p. 1247-1318.**
**186. Green. Economic Analysis. 2nd ed. New Jersey: Prentice Hall; 1997.**
**187. McClellan M, Staigner D. The quality of health care providers. .National Bureau of Economic Research Working Paper #7327.**
**188. Manheim LM, Sohn MW, Feinglass J, et al. Hospital vascular surgery volume and procedure mortality rates in California, 1982-1994. J Vasc Surg 1998;28(1):45-56.**

**189. Hannan EL, Popp AJ, Tranmer B, et al. Relationship between provider volume and mortality for carotid endarterectomies in New York state. Stroke 1998;29(11):2292-7.**
**190. Dudley RA, Johansen KL, Brand R, et al. Selective referral to high-volume hospitals: estimating potentially avoidable deaths. JAMA 2000;283(9):1159-66.**
**191. Ryan TJ, Bauman WB, Kennedy JW, et al. Guidelines for percutaneous transluminal coronary angioplasty. A report of the American Heart Association/American College of Cardiology Task Force on Assessment of Diagnostic and Therapeutic Cardiovascular Procedures (Committee on Percutaneous Transluminal Coronary Angioplasty). Circulation 1993;88(6):2987-3007.**
**192. Kimmel SE, Berlin JA, Laskey WK. The relationship between coronary angioplasty procedure volume and major complications. Jama 1995;274(14):1137-42.**
**193. Eagle KA, Guyton RA, Davidoff R, et al. ACC/AHA Guidelines for Coronary Artery Bypass Graft Surgery: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Revise the 1991 Guidelines for Coronary Artery Bypass Graft Surgery). American College of Cardiology/American Heart Association. J Am Coll Cardiol 1999;34(4):1262-347.**
**194. Hannan EL, Racz M, Kavey RE, et al. Pediatric cardiac surgery: the effect of hospital and surgeon volume on in-hospital mortality. Pediatrics 1998;101(6):963-9.**
**195. Hannan EL, Kilburn H, Jr., O'Donnell JF, et al. A longitudinal analysis of the relationship between in-hospital mortality in New York State and the volume of abdominal aortic aneurysm surgeries performed. Health Serv Res 1992;27(4):517-42.**
**196. Kazmers A, Jacobs L, Perkins A, et al. Abdominal aortic aneurysm repair in Veterans Affairs medical centers. J Vasc Surg 1996;23(2):191-200.**
**197. Pronovost PJ, Jenckes MW, Dorman T, et al. Organizational characteristics of intensive care units related to outcomes of abdominal aortic surgery. JAMA 1999;281(14):1310-7.**
**198. Patti MG, Corvera CU, Glasgow RE, et al. A hospital's annual rate of esophagectomy influences the operative mortality rate. J Gastrointest Surg 1998;2(2):186-92.**
**199. Glasgow RE, Mulvihill SJ. Hospital volume influences outcome in patients undergoing pancreatic resection for cancer. West J Med 1996;165(5):294-300.**
**200. Localio AR, Hamory BH, Sharp TJ, et al. Comparing hospital mortality in adult patients with pneumonia. A case study of statistical methods in a managed care program [see comments]. Ann Intern Med 1995;122(2):125-32.**
**201. Hannan EL, Kilburn H, Jr., Lindsey ML, et al. Clinical versus administrative data bases for CABG surgery. Does it matter? Med Care 1992;30(10):892-907.**
**202. Smith DW, Pine M, Bailey RC, et al. Using clinical variables to estimate the risk of patient mortality. Med Care 1991;29(11):1108-29.**
**203. Krakauer H, Bailey RC, Skellan KJ, et al. Evaluation of the HCFA model for the analysis of mortality following hospitalization. Health Serv Res 1992;27(3):317-35.**
**204. Landon B, Iezzoni LI, Ash AS, et al. Judging hospitals by severity-adjusted mortality rates: the case of CABG surgery. Inquiry 1996;33(2):155-66.**
**205. Pine M, Jones B, Lou YB. Laboratory values improve predictions of hospital mortality. Int J Qual Health Care 1998;10(6):491-501.**
**206. Sheiner LB, Easterling MJ, Mozes B, et al. Laboratory data predicts survival post hospitalization. J Clin Epidemiol 1991;44(12):1387-403.**

**207. Roos LL, Sharp SM, Cohen MM, et al. Risk adjustment in claims-based research: the search for efficient approaches. J Clin Epidemiol 1989;42(12):1193-206.**
**208. Averill RF, Muldoon JH, C. VJ, et al. The evolution of casemix measurement using diagnosis related groups (DRGs). .3M HIS working paper 5-98.**
**209. Burgess JF, Jr., Christiansen CL, Michalak SE, et al. Medical profiling: improving standards and risk adjustments using hierarchical models. J Health Econ 2000;19(3):291-309.**
**210. Donaldson MS, editor. Measuring the Quality of Health Care. A Statement by the National Roundtable on Health Care Quality. Washington, D.C.: National Academy Press; 1999.**
**211. Palmer RH, Reilly MC. Individual and institutional variables which may serve as indicators of quality of medical care. Med Care 1979;17(7):693-717.**
**212. Shortell SM, LoGerfo JP. Hospital medical staff organization and quality of care: results for myocardial infarction and appendectomy. Med Care 1981;19(10):1041-55.**
**213. Shortell SM, Hughes EF. The effects of regulation, competition, and ownership on mortality rates among hospital inpatients. N Engl J Med 1988;318(17):1100-7.**
**214. Hartz AJ, Krakauer H, Kuhn EM, et al. Hospital characteristics and mortality rates. N Engl J Med 1989;321(25):1720-5.**
**215. Kuhn EM, Hartz AJ, Gottlieb MS, et al. The relationship of hospital characteristics and the results of peer review in six large states. Med Care 1991;29(10):1028-38.**
**216. Brennan TA, Hebert LE, Laird NM, et al. Hospital characteristics associated with adverse events and substandard care. JAMA 1991;265(24):3265-9.**
**217. Keeler EB, Rubenstein LV, Kahn KL, et al. Hospital characteristics and quality of care. JAMA 1992;268(13):1709-14.**
**218. Ayanian JZ, Weissman JS, Chasan-Taber S, et al. Quality of care for two common illnesses in teaching and nonteaching hospitals. Health Aff (Millwood) 1998;17(6):194-205.**
**219. Taylor DH, Jr., Whellan DJ, Sloan FA. Effects of admission to a teaching hospital on the cost and quality of care for Medicare beneficiaries. N Engl J Med 1999;340(4):293-9.**
**220. Carson SS, Stocking C, Podsadecki T, et al. Effects of organizational change in the medical intensive care unit of a teaching hospital: a comparison of 'open' and 'closed' formats. JAMA 1996;276(4):322-8.**
**221. Collaborative systematic review of the randomised trials of organised inpatient (stroke unit) care after stroke. Stroke Unit Trialists' Collaboration. BMJ 1997;314(7088):1151-9.**
**222. Kaste M, Palomaki H, Sarna S. Where and how should elderly stroke patients be treated? A randomized trial. Stroke 1995;26(2):249-53.**
**223. Ronning OM, Guldvog B. Stroke unit versus general medical wards, II: neurological deficits and activities of daily living: a quasi-randomized controlled trial. Stroke 1998;29(3):586-90.**
**224. Pestotnik SL, Classen DC, Evans RS, et al. Implementing antibiotic practice guidelines through computer-assisted decision support: clinical and financial outcomes. Ann Intern Med 1996;124(10):884-90.**
**225. Bates DW, Leape LL, Cullen DJ, et al. Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. JAMA 1998;280(15):1311-6.**

**226. Evans RS, Pestotnik SL, Classen DC, et al. A computer-assisted management program for antibiotics and other antiinfective agents. N Engl J Med 1998;338(4):232-8.**
**227. Hughes RG, Hunt SS, Luft HS. Effects of surgeon volume and hospital volume on quality of care in hospitals. Med Care 1987;25(6):489-503.**
**228. Showstack JA, Rosenfeld KE, Garnick DW, et al. Association of volume with outcome of coronary artery bypass graft surgery. Scheduled vs nonscheduled operations. JAMA 1987;257(6):785-9.**
**229. Kelly JV. What do mortality studies reveal about hospital volume, teaching status, and ownership? Int J Technol Assess Health Care 1990;6(2):239-52.**
**230. Farley DE, Ozminkowski RJ. Volume-outcome relationships and in-hospital mortality: the effect of changes in volume over time. Med Care 1992;30(1):77-94.**
**231. Byrne MG, Giacomini JC. Longitudinal comparison of outcomes for coronary angioplasty at low, medium and high volume hospitals in California for years 1988-1992. Circulation 1995;92(8 SUPPL.):I47.**
**232. Byrne MG, Reitz BA. Longitudinal comparison of outcomes for coronary artery bypass surgery at low, medium and high volume hospitals in California for years 1985-1991. Circulation 1995;92(8 SUPPL.):I47.**
**233. Hannan EL, Siu AL, Kumar D, et al. The decline in coronary artery bypass graft surgery mortality in New York State. The role of surgeon volume. JAMA 1995;273(3):209-13.**
**234. Lavernia CJ, Guzman JF. Relationship of surgical volume to short-term mortality, morbidity, and hospital charges in arthroplasty. J Arthroplasty 1995;10(2):133-40.**
**235. Phillips KA, Luft HS, Ritchie JL. The association of hospital volumes of percutaneous transluminal coronary angioplasty with adverse outcomes, length of stay, and charges in California. Med Care 1995;33(5):502-14.**
**236. Taylor HD, Dennis DA, Crane HS. Relationship between mortality rates and hospital patient volume for Medicare patients undergoing major orthopaedic surgery of the hip, knee, spine, and femur. J Arthroplasty 1997;12(3):235-42.**
**237. McGrath PD, Wennberg DE, Malenka DJ, et al. Operator volume and outcomes in 12,998 percutaneous coronary interventions. Northern New England Cardiovascular Disease Study Group. J Am Coll Cardiol 1998;31(3):570-6.**
**238. Glasgow RE, Showstack JA, Katz PP, et al. The relationship between hospital volume and outcomes of hepatic resection for hepatocellular carcinoma. Arch Surg 1999;134(1):30-5.**
**239. Luft HS, Hunt SS, Maerki SC. The volume-outcome relationship: practice-makes-perfect or selective- referral patterns? Health Serv Res 1987;22(2):157-82.**
**240. Hamilton BH, Ho V. Does practice make perfect? Examining the relationship between hospital surgical volume and outcomes for hip fracture patients in Quebec. Med Care 1998;36(6):892-903.**
**241. Luft HS, Garnick DW, Mark DH, et al. Does quality influence choice of hospital? JAMA 1990;263(21):2899-906.**
**242. Hamilton BH, Hamilton VH. Estimating surgical volume--outcome relationships applying survival models: accounting for frailty and hospital fixed effects. Health Econ 1997;6(4):383-95.**
**243. Tepas JJ, 3rd, Patel JC, DiScala C, et al. Relationship of trauma patient volume to outcome experience: can a relationship be defined? J Trauma 1998;44(5):827-30.**

**244. Phillips KA, Luft HS. The policy implications of using hospital and physician volumes as "indicators" of quality of care in a changing health care environment. Int J Qual Health Care 1997;9(5):341-8.**
**245. Finlayson SR, Birkmeyer JD, Tosteson AN, et al. Patient preferences for location of care: implications for regionalization. Med Care 1999;37(2):204-9.**
**246. Roos NP, Black CD, Roos LL, et al. A population-based approach to monitoring adverse outcomes of medical care. Med Care 1995;33(2):127-38.**
**247. North American Symptomatic Carotid Endarterectomy Trial. Methods, patient characteristics, and progress. Stroke 1991;22(6):711-20.**
**248. Endarterectomy for asymptomatic carotid artery stenosis. Executive Committee for the Asymptomatic Carotid Atherosclerosis Study. JAMA 1995;273(18):1421-8.**
**249. Holloway RG, Jr., Witter DM, Jr., Mushlin AI, et al. Carotid endarterectomy trends in the patterns and outcomes of care at academic medical centers, 1990 through 1995. Arch Neurol 1998;55(1):25-32.**
**250. Hsia DC, Moscoe LM, Krushat WM. Epidemiology of carotid endarterectomy among Medicare beneficiaries: 1985-1996 update. Stroke 1998;29(2):346-50.**
**251. Tu JV, Hannan EL, Anderson GM, et al. The fall and rise of carotid endarterectomy in the United States and Canada. N Engl J Med 1998;339(20):1441-7.**
**252. Barnett HJ, Meldrum HE, Eliasziw M. The dilemma of surgical treatment for patients with asymptomatic carotid disease. Ann Intern Med 1995;123(9):723-5.**
**253. Gould DA, Birkmeyer JD. Efficacy versus effectiveness of carotid endarterectomy. Eff Clin Pract 1999;2(1):30-6.**
**254. Benavente O, Moher D, Pham B. Carotid endarterectomy for asymptomatic carotid stenosis: a meta- analysis. BMJ 1998;317(7171):1477-80.**
**255. Warlow C. Carotid endarterectomy for asymptomatic carotid stenosis. Better data, but the case is still not convincing. BMJ 1998;317(7171):1468.**
**256. Mayo SW, Eldrup-Jorgensen J, Lucas FL, et al. Carotid endarterectomy after NASCET and ACAS: a statewide study. North American Symptomatic Carotid Endarterectomy Trial. Asymptomatic Carotid Artery Stenosis Study. J Vasc Surg 1998;27(6):1017-22.**
**257. Hallett JW, Jr., Pietropaoli JA, Jr., Ilstrup DM, et al. Comparison of North American Symptomatic Carotid Endarterectomy Trial and population-based outcomes for carotid endarterectomy. J Vasc Surg 1998;27(5):845-50.**
**258. Wennberg DE, Lucas FL, Birkmeyer JD, et al. Variation in carotid endarterectomy mortality in the Medicare population: trial hospitals, volume, and patient characteristics. JAMA 1998;279(16):1278-81.**
**259. Huber TS, Wheeler KG, Cuddeback JK, et al. Effect of the Asymptomatic Carotid Atherosclerosis Study on carotid endarterectomy in Florida. Stroke 1998;29(6):1099-105.**
**260. Chaturvedi S. Is the increasing volume of carotid endarterectomy justified? [letter]. Stroke 1998;29(11):2444-6.**
**261. Keller RB, Atlas SJ, Soule DN, et al. Relationship between rates and outcomes of operative treatment for lumbar disc herniation and spinal stenosis. J Bone Joint Surg Am 1999;81(6):752-62.**
**262. Leape LL, Hilborne LH, Park RE, et al. The appropriateness of use of coronary artery bypass graft surgery in New York State. Jama 1993;269(6):753-60.**

**263. Bernstein SJ, McGlynn EA, Siu AL, et al. The appropriateness of hysterectomy. A comparison of care in seven health plans. Health Maintenance Organization Quality of Care Consortium [see comments]. Jama 1993;269(18):2398-402.**
**264. Bernstein SJ, Hilborne LH, Leape LL, et al. The appropriateness of use of coronary angiography in New York State. Jama 1993;269(6):766-9.**
**265. Soumerai SB, Ross-Degnan D, Fortess EE, et al. Determinants of change in Medicaid pharmaceutical cost sharing: does evidence affect policy? Milbank Q 1997;75(1):11-34.**
**266. Chassin MR. Appropriate use of carotid endarterectomy [editorial; comment] [see comments]. N Engl J Med 1998;339(20):1468-71.**
**267. Kravitz RL, Laouri M, Kahan JP, et al. Validity of criteria used for detecting underuse of coronary revascularization. JAMA 1995;274(8):632-8.**
**268. Chassin MR, Brook RH, Park RE, et al. Variations in the use of medical and surgical services by the Medicare population. N Engl J Med 1986;314(5):285-90.**
**269. Shekelle PG, Kahan JP, Bernstein SJ, et al. The reproducibility of a method to identify the overuse and underuse of medical procedures. N Engl J Med 1998;338(26):1888-95.**
**270. Asaph JW, Janoff K, Wayson K, et al. Carotid endarterectomy in a community hospital: a change in physicians' practice patterns. Am J Surg 1991;161(5):616-8.**
**271. Brott TG, Labutta RJ, Kempczinski RF. Changing patterns in the practice of carotid endarterectomy in a large metropolitan area. Jama 1986;255(19):2609-12.**
**272. Geller SE, Burns LR, Brailer DJ. The impact of nonclinical factors on practice variations: the case of hysterectomies. Health Serv Res 1996;30(6):729-50.**
**273. Marshak JD, Cooperman BS, Fried WB, et al. Prognostic indicators for successful vaginal birth after cesarean delivery. Obstet Gynecol 2000;95(4 Suppl 1):S38.**
**274. Brook RHI. The RAND/UCLA appropriateness method. In: McCormick KA, Moore SR, Siegel RA, editors. Clinical Practice Guideline Development: Methodology Perspectives, AHCPR Pub. No. 95-0009. Rockville, MD: Public Health Service; 1994. p. 59-70.**
**275. Billings J, Zeitel L, Lukomnik J, et al. Impact of socioeconomic status on hospital use in New York City. Health Aff (Millwood) 1993;12(1):162-73.**
**276. Weissman JS, Gatsonis C, Epstein AM. Rates of avoidable hospitalization by insurance status in Massachusetts and Maryland. Jama 1992;268(17):2388-94.**
**277. Gadomski A, Jenkins P, Nichols M. Impact of a Medicaid primary care provider and preventive care on pediatric hospitalization. Pediatrics 1998;101(3):E1.**
**278. Pediatric Ambulatory Care Sensitive Conditions in South Carolina. Columbia: South Carolina State Budget and Control Board, Office of Research and Statistics; 1995.**
**279. Schreiber S, Zielinski T. The meaning of ambulatory care sensitive admissions: urban and rural perspectives. J Rural Health 1997;13(4):276-84.**
**280. Culler SD, Parchman ML, Przybylski M. Factors related to potentially preventable hospitalizations among the elderly. Med Care 1998;36(6):804-17.**
**281. Pappas G, Hadden WC, Kozak LJ, et al. Potentially avoidable hospitalizations: inequalities in rates between US socioeconomic groups. Am J Public Health 1997;87(5):811-6.**
**282. Blustein J, Hanson K, Shea S. Preventable hospitalizations and socioeconomic status. Health Aff (Millwood) 1998;17(2):177-89.**

**283. Billings J, Anderson GM, Newman LS. Recent findings on preventable hospitalizations . Health Aff (Millwood) 1996;15(3):239-49.**
**284. Bindman AB, Grumbach K, Osmond D, et al. Preventable hospitalizations and access to health care. JAMA 1995;274(4):305-11.**
**285. Parchman ML, Culler SD. Preventable hospitalizations in primary care shortage areas. An analysis of vulnerable Medicare beneficiaries. Arch Fam Med 1999;8(6):487-91.**
**286. Parchman ML, Culler S. Primary care physicians and avoidable hospitalizations . J Fam Pract 1994;39(2):123-8.**
**287. McConnochie KM, Roghmann KJ, Liptak GS. Socioeconomic variation in discretionary and mandatory hospitalization of infants: an ecologic analysis . Pediatrics 1997;99(6):774-84.**
**288. Josephson GW, Karcz A. The impact of physician economic incentives on admission rates of patients with ambulatory sensitive conditions: an analysis comparing two managed care structures and indemnity insurance. Am J Manag Care 1997;3(1):49-56.**
**289. Chassin MR, Park RE, Lohr KN, et al. Differences among hospitals in Medicare patient mortality. Health Serv Res 1989;24(1):1-31.**
**290. Beguin C, France FR, Ninane J. Systematic analysis of in-patients' circumstances and causes of death: a tool to improve quality of care. Int J Qual Health Care 1997;9(6):427-33.**
**291. Dubois RW, Brook RH. Preventable deaths: who, how often, and why? Ann Intern Med 1988;109(7):582-9.**
**292. Rosenthal GE, Quinn L, Harper DL. Declines in hospital mortality associated with a regional initiative to measure hospital performance. Am J Med Qual 1997;12(2):103-12.**
**293. Nationwide inpatient sample (NIS). ,HCUPnet, Healthcare Cost and Utilization Project**
**Agency for Healthcare Reserach and Quality, Rockville, MD.**
**294. Pearce WH, Parker MA, Feinglass J, et al. The importance of surgeon volume and training in outcomes for vascular surgical procedures. J Vasc Surg 1999;29(5):768-76.**
**295. Sollano JA, Gelijns AC, Moskowitz AJ, et al. Volume-outcome relationships in cardiovascular operations: New York State, 1990-1995. J Thorac Cardiovasc Surg 1999;117(3):419-28.**
**296. Dardik A, Lin JW, Gordon TA, et al. Results of elective abdominal aortic aneurysm repair in the 1990s: A population-based analysis of 2335 cases. J Vasc Surg 1999;30(6):985-95.**
**297. Owings MF, Kozak LJ. Ambulatory and inpatient procedures in the United States, 1996. Vital Health Stat 13 1998(139):1-119.**
**298. Khuri SF, Daley J, Henderson W, et al. Relation of surgical volume to outcome in eight common operations: results from the VA National Surgical Quality Improvement Program. Ann Surg 1999;230(3):414-29.**
**299. Rutledge R, Oller DW, Meyer AA, et al. A statewide, population-based time-series analysis of the outcome of ruptured abdominal aortic aneurysm. Ann Surg 1996;223(5):492-502.**
**300. Katz DJ, Stanley JC, Zelenock GB. Operative mortality rates for intact and ruptured abdominal aortic aneurysms in Michigan: an eleven-year statewide experience. J Vasc Surg 1994;19(5):804-15; discussion 816-7.**

**301. Wen SW, Simunovic M, Williams JI, et al. Hospital volume, calendar age, and short term outcomes in patients undergoing repair of abdominal aortic aneurysms: the Ontario experience, 1988-92. J Epidemiol Community Health 1996;50(2):207-13.**
**302. Amundsen S, Skjaerven R, Trippestad A, et al. Abdominal aortic aneurysms. Is there an association between surgical volume, surgical experience, hospital type and operative mortality? Members of the Norwegian Abdominal Aortic Aneurysm Trial. Acta Chir Scand 1990;156(4):323-7; discussion 327-8.**
**303. Kantonen I, Lepantalo M, Brommels M, et al. Mortality in ruptured abdominal aortic aneurysms. The Finnvasc Study Group. Eur J Vasc Endovasc Surg 1999;17(3):208-12.**
**304. Pacific Business Group on Health Web Page,. ,Pacific Business Group on Health.**
**305. ,The Center for Medical Consumers.**
**306. The Asymptomatic Carotid Endarterectomy Trial Collaborators. Study design for randomized prospective trial of carotid endarterectomy for asymtomatic atherosclerosis. Stroke 1989;20:844-849.**
**307. Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis. North American Symptomatic Carotid Endarterectomy Trial Collaborators. N Engl J Med 1991;325(7):445-53.**
**308. Biller J, Feinberg WM, Castaldo JE, et al. Guidelines for carotid endarterectomy: a statement for healthcare professionals from a Special Writing Group of the Stroke Council, American Heart Association. Circulation 1998;97(5):501-9.**
**309. Owings MF, Lawrence L. Detailed diagnoses and procedures, National Hospital Discharge Survey, 1997. Vital Health Stat 13 1999(145):1-157.**
**310. Karp HR, Flanders WD, Shipp CC, et al. Carotid endarterectomy among Medicare beneficiaries: a statewide evaluation of appropriateness and outcome. Stroke 1998;29(1):46-52.**
**311. Fisher ES, Malenka DJ, Solomon NA, et al. Risk of carotid endarterectomy in the elderly. Am J Public Health 1989;79(12):1617-20.**
**312. Edwards WH, Morris JA, Jr., Jenkins JM, et al. Evaluating quality, cost-effective health care. Vascular database predicated on hospital discharge abstracts. Ann Surg 1991;213(5):433-8.**
**313. Kantonen I, Lepantalo M, Salenius JP, et al. Mortality in abdominal aortic aneurysm surgery--the effect of hospital volume, patient mix and surgeon's case load. Eur J Vasc Endovasc Surg 1997;14(5):375-9.**
**314. Barnett HJ, Taylor DW, Eliasziw M, et al. Benefit of carotid endarterectomy in patients with symptomatic moderate or severe stenosis. North American Symptomatic Carotid Endarterectomy Trial Collaborators. N Engl J Med 1998;339(20):1415-25.**
**315. Crawford FA, Jr., Anderson RP, Clark RE, et al. Volume requirements for cardiac surgery credentialing: a critical examination. The Ad Hoc Committee on Cardiac Surgery Credentialing of The Society of Thoracic Surgeons [see comments]. Ann Thorac Surg 1996;61(1):12-6.**
**316. Kozak LJ, Lawrence L. National hospital discharge survey: annual summary, 1997. Vital Health Stat 13 1999(144):i-iv, 1-46.**
**317. Gray D, Hampton JR, Bernstein SJ, et al. Audit of coronary angiography and bypass surgery [see comments]. Lancet 1990;335(8701):1317-20.**

**318. Choosing a Hospital for Surgery - Coronary Artery Bypass Graft Surgery. ,Pacific Business Group on Health.**
**319. Sowden AJ, Deeks JJ, Sheldon TA. Volume and outcome in coronary artery bypass graft surgery: true association or artefact? [see comments]. Bmj 1995;311(6998):151-5.**
**320. Shroyer AL, Marshall G, Warner BA, et al. No continuous relationship between Veterans Affairs hospital coronary artery bypass grafting surgical volume and operative mortality. Ann Thorac Surg 1996;61(1):17-20.**
**321. Kelly JV, Hellinger FJ. Physician and hospital factors associated with mortality of surgical patients. Med Care 1986;24(9):785-800.**
**322. Gordon TA, Bowman HM, Bass EB, et al. Complex gastrointestinal surgery: impact of provider experience on clinical and economic outcomes. J Am Coll Surg 1999;189(1):46-56.**
**323. Matthews HR, Powell DJ, McConkey CC. Effect of surgical experience on the results of resection for oesophageal carcinoma. Br J Surg 1986;73(8):621-3.**
**324. Lieberman MD, Kilburn H, Lindsey M, et al. Relation of perioperative deaths to hospital volume among patients undergoing pancreatic resection for malignancy. Ann Surg 1995;222(5):638-45.**
**325. Sosa JA, Bowman HM, Gordon TA, et al. Importance of hospital volume in the overall management of pancreatic cancer. Ann Surg 1998;228(3):429-38.**
**326. Simunovic M, To T, Theriault M, et al. Relation between hospital surgical volume and outcome for pancreatic resection for neoplasm in a publicly funded health care system [see comments]. Cmaj 1999;160(5):643-8.**
**327. Neoptolemos JP, Russell RC, Bramhall S, et al. Low mortality following resection for pancreatic and periampullary tumours in 1026 patients: UK survey of specialist pancreatic units. UK Pancreatic Cancer Group [see comments]. Br J Surg 1997;84(10):1370-6.**
**328. Birkmeyer JD, Finlayson SR, Tosteson AN, et al. Effect of hospital volume on in-hospital mortality with pancreaticoduodenectomy. Surgery 1999;125(3):250-6.**
**329. Wade TP, Halaby IA, Stapleton DR, et al. Population-based analysis of treatment of pancreatic cancer and Whipple resection: Department of Defense hospitals, 1989-1994. Surgery 1996;120(4):680-5; discussion 686-7.**
**330. Edge SB, Schmieg RE, Jr., Rosenlof LK, et al. Pancreas cancer resection outcome in American University centers in 1989-1990. Cancer 1993;71(11):3502-8.**
**331. Gordon TA, Bowman HM, Tielsch JM, et al. Statewide regionalization of pancreaticoduodenectomy and its effect on in-hospital mortality. Ann Surg 1998;228(1):71-8.**
**332. Jenkins KJ, Newburger JW, Lock JE, et al. In-hospital mortality for surgical repair of congenital heart defects: preliminary observations of variation by hospital caseload. Pediatrics 1995;95(3):323-30.**
**333. Hirshfeld JW, Jr., Ellis SG, Faxon DP. Recommendations for the assessment and maintenance of proficiency in coronary interventional procedures: Statement of the American College of Cardiology. J Am Coll Cardiol 1998;31(3):722-43.**
**334. Ryan TJ, Klocke FJ, Reynolds WA. Clinical competence in percutaneous transluminal coronary angioplasty. A statement for physicians from the ACP/ACC/AHA Task Force on Clinical Privileges in Cardiology. J Am Coll Cardiol 1990;15(7):1469-74.**

**335. Ritchie JL, Phillips KA, Luft HS. Coronary angioplasty. Statewide experience in California. Circulation 1993;88(6):2735-43.**
**336. Shook TL, Sun GW, Burstein S, et al. Comparison of percutaneous transluminal coronary angioplasty outcome and hospital costs for low-volume and high-volume operators. Am J Cardiol 1996;77(5):331-6.**
**337. Jollis JG, Peterson ED, DeLong ER, et al. The relation between the volume of coronary angioplasty procedures at hospitals treating Medicare beneficiaries and short-term mortality. N Engl J Med 1994;331(24):1625-9.**
**338. Malenka DJ, McGrath PD, Wennberg DE, et al. The relationship between operator volume and outcomes after percutaneous coronary interventions in high volume hospitals in 1994- 1996: the northern New England experience. Northern New England Cardiovascular Disease Study Group. J Am Coll Cardiol 1999;34(5):1471-80.**
**339. Menard MK. Cesarean delivery rates in the United States. The 1990s. Obstet Gynecol Clin North Am 1999;26(2):275-86.**
**340. Parrish KM, Holt VL, Easterling TR, et al. Effect of changes in maternal age, parity, and birth weight distribution on primary cesarean delivery rates. JAMA 1994;271(6):443-7.**
**341. Cesarean childbirth: report of the NICHD task force on cesarean childbirth. Bethesda: United States Department of Health and Human Services National Institutes of Health; 1981. Report No.: DHHS publicaton no. [NIH] 82-2067.**
**342. Elixhauser A, Steiner CA. Most Common Diagnoses and Procedures in U.S. Community Hospitals 1996**
**Summary**
**HCUP Research Note. ,Agency for Health Care Policy and Research Rockville MD.**
**343. ACOG Committee CoOP. Vaginal Delivery After a Previous Cesarean Birth. Washington, D. C.: The American College of Obstetricians and Gynecologists; 1994 October. Report No.: 143.**
**344. Goyert GL, Bottoms SF, Treadwell MC, et al. The physician factor in cesarean birth rates [see comments]. N Engl J Med 1989;320(11):706-9.**
**345. Berkowitz GS, Fiarman GS, Mojica MA, et al. Effect of physician characteristics on the cesarean birth rate [see comments]. Am J Obstet Gynecol 1989;161(1):146-9.**
**346. DeMott RK, Sandmire HF. The Green Bay cesarean section study. I. The physician factor as a determinant of cesarean birth rates. Am J Obstet Gynecol 1990;162(6):1593-9; discussion 1599-602.**
**347. Burns LR, Geller SE, Wholey DR. The effect of physician factors on the cesarean section decision. Med Care 1995;33(4):365-82.**
**348. Guillemette J, Fraser WD. Differences between obstetricians in caesarean section rates and the management of labour. Br J Obstet Gynaecol 1992;99(2):105-8.**
**349. Tussing AD, Wojtowycz MA. The cesarean decision in New York State, 1986. Economic and noneconomic aspects. Med Care 1992;30(6):529-40.**
**350. Jessee WF, Nickerson CW, Grant WS. Assessing medical practices through PSRO cooperative studies, an evaluation of Cesarean births in nine PSRO areas. Med Care 1982;20(1):75-84.**
**351. Placek PJ, Taffel SM. Recent patterns in cesarean delivery in the United States. Obstet Gynecol Clin North Am 1988;15(4):607-27.**

**352. Gould JB, Davey B, Stafford RS. Socioeconomic differences in rates of cesarean section [see comments]. N Engl J Med 1989;321(4):233-9.**
**353. Stafford RS. Cesarean section use and source of payment: an analysis of California hospital discharge abstracts. Am J Public Health 1990;80(3):313-5.**
**354. Stafford RS. The impact of nonclinical factors on repeat cesarean section [see comments]. Jama 1991;265(1):59-63.**
**355. Haas JS, Udvarhelyi S, Epstein AM. The effect of health coverage for uninsured pregnant women on maternal health and the use of cesarean section [see comments]. Jama 1993;270(1):61-4.**
**356. Stafford RS, Sullivan SD, Gardner LB. Trends in cesarean section use in California, 1983 to 1990. Am J Obstet Gynecol 1993;168(4):1297-302.**
**357. Keller EB, Park RE, Bell RM, et al. Adjusting cesarean delivery rates for case mix. Health Serv Res 1997;43(4):511-528.**
**358. Aron DC, Harper DL, Shepardson LB, et al. Impact of risk-adjusting cesarean delivery rates when reporting hospital performance. JAMA 1998;279(24):1968-72.**
**359. Bailit JL, Dooley SL, Peaceman AN. Risk adjustment for interhospital comparison of primary cesarean rates. Obstet Gynecol 1999;93(6):1025-30.**
**360. Glantz JC. Cesarean delivery risk adjustment for regional interhospital comparisons. American Journal of Obstetrics and Gynecology 1999;181(6):1425-31.**
**361. Gregory KD, Ramicone E, Chan L, et al. Cesarean deliveries for medicaid patients: a comparison in public and private hospitals in Los Angeles county. Am J Obstet Gynecol 1999;180(5):1177-84.**
**362. Anderson GM, Lomas J. Determinants of the increasing cesarean birth rate. Ontario data 1979 to 1982. N Engl J Med 1984;311(14):887-92.**
**363. Henry OA, Gregory KD, Hobel CJ, et al. Using ICD-9 codes to identify indications for primary and repeat cesarean sections: agreement with clinical records. American Journal of Public Health 1995;85(8 Pt 1):1143-6.**
**364. Gifford DS, Morton SC, Fiske M, et al. A meta-analysis of infant outcomes after breech delivery. Obstet Gynecol 1995;85(6):1047-54.**
**365. Sachs BP, Kobelin C, Castro MA, et al. The risks of lowering the cesarean-delivery rate. N Engl J Med 1999;340(1):54-7.**
**366. Lagrew DC, Jr., Adashek JA. Lowering the cesarean section rate in a private hospital: comparison of individual physicians' rates, risk factors, and outcomes. Am J Obstet Gynecol 1998;178(6):1207-14.**
**367. Sandmire HF, DeMott RK. The Green Bay cesarean section study. IV. The physician factor as a determinant of cesarean birth rates for the large fetus. Am J Obstet Gynecol 1996;174(5):1557-64.**
**368. Technical Report: Report Card Pilot Project. Washington, D. C.: National Committee for Quality Assurance; 1995.**
**369. Maryland Quality Indicators Project. ,Maryland Hospital Association.**
**370. University Hospital Consortium. ,University Hospital Consortium.**
**371. Florida Agency for Health Care Administration Web Page. ,Florida Agency for Health Care Administration.**
**372. Greater New York Hospital Association. .**
**373. Michigan Hospital Association Web Page. ,Michigan Hospital Association.**
**374. Cleveland Health Quality Choice. .**

**375. Virginia Health Information Web Page. ,Virginia Health Information.**
**376. Washington State Community Health Information Partnership. ,CHMIS.**
**377. HealthGrades.com. ,HealthGrades.com.**
**378. Kelly H. Under what circumstances is it advisable to remove the veriform appendix when the abdomen is opened for other reasons. JAMA 1902;39:1014-1021.**
**379. O'Malley VP, Finch DR, Powley PH. Wound sepsis after cholecystectomy. Influence of incidental appendectomy. J Clin Gastroenterol 1986;8(4):435-7.**
**380. Strom PR, Turkleson ML, Stone HH. Safety of incidental appendectomy. Am J Surg 1983;145(6):819-22.**
**381. Ikard RW. Prospective analysis of the effect of incidental appendectomy on infection rate after cholecystectomy. South Med J 1987;80(3):292-5.**
**382. Pollock AV, Evans M. Wound sepsis after cholecystectomy: effect of incidental appendicectomy. Br Med J 1977;1(6052):20-2.**
**383. Warren JL, Penberthy LT, Addiss DG, et al. Appendectomy incidental to cholecystectomy among elderly Medicare beneficiaries. Surg Gynecol Obstet 1993;177(3):288-94.**
**384. Wen SW, Hernandez R, Naylor CD. Pitfalls in nonrandomized outcomes studies. The case of incidental appendectomy with open cholecystectomy. Jama 1995;274(21):1687-91.**
**385. Andrew MH, Roty AR, Jr. Incidental appendectomy with cholecystectomy: is the increased risk justified? Am Surg 1987;53(10):553-7.**
**386. Fisher KS, Ross DS. Guidelines for therapeutic decision in incidental appendectomy. Surg Gynecol Obstet 1990;171(1):95-8.**
**387. Snyder TE, Selanders JR. Incidental appendectomy--yes or no? A retrospective case study and review of the literature. Infect Dis Obstet Gynecol 1998;6(1):30-7.**
**388. Wolff BG. Current status of incidental surgery. Dis Colon Rectum 1995;38(4):435-41.**
**389. Nockerts SR, Detmer DE, Fryback DG. Incidental appendectomy in the elderly? No. Surgery 1980;88(2):301-6.**
**390. Neulander EZ, Hawke CK, Soloway MS. Incidental appendectomy during radical cystectomy: an interdepartmental survey and review of the literature. Urology 2000;56(2):241-4.**
**391. Hill JA, Miranda AA, Keim SG, et al. Value of right-sided cardiac catheterization in patients undergoing left-sided cardiac catheterization for evaluation of coronary artery disease. Am J Cardiol 1990;65(9):590-3.**
**392. Shanes JG, Stein MA, Dierenfeldt BJ, et al. The value of routine right heart catheterization in patients undergoing coronary arteriography. Am Heart J 1987;113(5):1261-3.**
**393. Friedman HS. Right-heart catheterization in coronary artery disease. Angiology 1978;29(12):878-87.**
**394. Barron JT, Ruggie N, Uretz E, et al. Findings on routine right heart catheterization in patients with suspected coronary artery disease. Am Heart J 1988;115(6):1193-8.**
**395. Pepine CJ, Allen HD, Bashore TM, et al. ACC/AHA guidelines for cardiac catheterization and cardiac catheterization laboratories. American College of Cardiology/American Heart Association Ad Hoc Task Force on Cardiac Catheterization. Circulation 1991;84(5):2213-47.**

**396. Quality Resume. Health Care Finiancing Administration's Medicare Quality of Care Report of Surveillance Measures. .**
**397. Hall M, Popovic J. 1998 summary: National Hospital Discharge Survey. Advance Data from Vital and Health Statistics 2000;316.**
**398. Malone ML, Bajwa TK, Battiola RJ, et al. Variation among cardiologists in the utilization of right heart catheterization at time of coronary angiography [see comments]. Cathet Cardiovasc Diagn 1996;37(2):125-30.**
**399. Medicare Quality of Care Report of Surveillance Measures. ,Health Care Financing Administration.**
**400. American Health Quality Association. A Pillar of Quality:The Medicare Peer Review Organization/Quality Improvement Organization Program. .**
**401. Bing ML, Abel RL, Lee LJ, et al. Medical necessity for right heart catheterization. Tex Heart Inst J 1997;24(2):109-13.**
**402. Fortune GJ, Schiffel F, Jr., Elder S. MPCRF: the Right Heart Catheterization Cooperative project. Mo Med 1996;93(10):657-61.**
**403. Gold JA. Decreasing the rate of bilateral cardiac catheterization. Wis Med J 1995;94(10):569-70.**
**404. Malach M, Imperato PJ, Nenner RP, et al. Impact of an educational program on bilateral heart catheterization practice patterns. Am J Med Qual 1998;13(4):213-22.**
**405. Imperato PJ, Malach M, Nenner RP, et al. Concurrent improvements in ambulatory cardiac catheterization practices following inpatient interventions. J Ambulatory Care Manage 1999;22(2):1-8.**
**406. US Dept. of Health and Human Services: Public Health Service National Institutes of Heatl. Repeat Cesarean Brith. In: Cesarean Childbirth. Wahington D.C.: US Government Printing Office; 1991. p. 351-74.**
**407. Phelan JP, Clark SL, Diaz F, et al. Vaginal birth after cesarean. Am J Obstet Gynecol 1987;157(6):1510-5.**
**408. Rosen MG, Dickinson JC, Westhoff CL. Vaginal birth after cesarean: a meta-analysis of morbidity and mortality. Obstet Gynecol 1991;77(3):465-70.**
**409. Flamm BL, Newman LA, Thomas SJ, et al. Vaginal birth after cesarean delivery: results of a 5-year multicenter collaborative study. Obstet Gynecol 1990;76(5 Pt 1):750-4.**
**410. Flamm BL, Goings JR, Liu Y, et al. Elective repeat cesarean delivery versus trial of labor: a prospective multicenter study. Obstet Gynecol 1994;83(6):927-32.**
**411. Cowan RK, Kinch RA, Ellis B, et al. Trial of labor following cesarean delivery. Obstet Gynecol 1994;83(6):933-6.**
**412. Miller DA, Diaz FG, Paul RH. Vaginal birth after cesarean: a 10-year experience. Obstet Gynecol 1994;84(2):255-8.**
**413. Joseph GF, Jr., Stedman CM, Robichaux AG. Vaginal birth after cesarean section: the impact of patient resistance to a trial of labor. Am J Obstet Gynecol 1991;164(6 Pt 1):1441-4; discussion 1444-7.**
**414. Kline J, Arias F. Analysis of factors determining the selection of repeated cesarean section or trial of labor in patients with histories of prior cesarean delivery. J Reprod Med 1993;38(4):289-92.**
**415. Miller ES, Partezana J, Montgomery RL. Vaginal birth after cesarean: a 5-year experience in a family practice residency program. J Am Board Fam Pract 1995;8(5):357-60.**

**416. Roberts RG, Bell HS, Wall EM, et al. Trial of labor or repeated cesarean section. The woman's choice . Arch Fam Med 1997;6(2):120-5.**
**417. Al-Mufti R, McCarthy A, Fisk NM. Obstetricians' personal choice and mode of delivery [letter] [see comments]. Lancet 1996;347(9000):544.**
**418. Clark SL, Scott JR, Porter TF, et al. Is vaginal birth after cesarean less expensive than repeat cesarean delivery? Am J Obstet Gynecol 2000;182(3):599-602.**
**419. Jones RO, Nagashima AW, Hartnett-Goodman MM, et al. Rupture of low transverse cesarean scars during trial of labor [see comments]. Obstet Gynecol 1991;77(6):815-7.**
**420. Scott JR. Mandatory trial of labor after cesarean delivery: an alternative viewpoint [see comments]. Obstet Gynecol 1991;77(6):811-4.**
**421. McMahon MJ, Luther ER, Bowes WA, Jr., et al. Comparison of a trial of labor with an elective second cesarean section. N Engl J Med 1996;335(10):689-95.**
**422. Hook B, Kiwi R, Amini SB, et al. Neonatal morbidity after elective repeat cesarean section and trial of labor. Pediatrics 1997;100(3 Pt 1):348-53.**
**423. Caughey AB, Shipp TD, Repke JT, et al. Trial of labor after cesarean delivery: the effect of previous vaginal delivery. Am J Obstet Gynecol 1998;179(4):938-41.**
**424. Gregory KD, Korst LM, Cane P, et al. Vaginal birth after cesarean and uterine rupture rates in California. Obstet Gynecol 1999;94(6):985-9.**
**425. Use of hospital discharge data to monitor uterine rupture -- Massachusetts, 1990-1997. MMWR Morb Mortal Wkly Rep 2000;49(12):245-8.**
**426. Obara H, Minakami H, Koike T, et al. Vaginal birth after cesarean delivery: results in 310 pregnancies. J Obstet Gynaecol Res 1998;24(2):129-34.**
**427. Rageth JC, Juzi C, Grossenbacher H. Delivery after previous cesarean: a risk evaluation. Swiss Working Group of Obstetric and Gynecologic Institutions. Obstet Gynecol 1999;93(3):332-7.**
**428. What is the right number of caesarean sections? . Lancet 1997;349(9055):815.**
**429. Green DC, Moore JM, Adams MM, et al. Are we underestimating rates of vaginal birth after previous cesarean birth? The validity of delivery methods from birth certificates. Am J Epidemiol 1998;147(6):581-6.**
**430. King DE, Lahiri K. Socioeconomic factors and the odds of vaginal birth after cesarean delivery . JAMA 1994;272(7):524-9.**
**431. Pollard JK, Capeless EL. Cesarean deliveries at a university hospital: analysis of rates and indications. Am J Perinatol 1997;14(5):245-8.**
**432. Lieberman E, Lang JM, Heffner LJ, et al. Assessing the role of case mix in cesarean delivery rates. Obstet Gynecol 1998;92(1):1-7.**
**433. Clark SL, Xu W, Porter TF, et al. Institutional influences on the primary cesarean section rate in Utah, 1992 to 1995. Am J Obstet Gynecol 1998;179(4):841-5.**
**434. Pasternak DP, Pine M, Nolan K, et al. Risk-adjusted measurement of primary cesarean sections: reliable assessment of the quality of obstetrical services. Qual Manag Health Care 1999;8(1):47-54.**
**435. Clarke SC, Taffel SM. State variation in rates of cesarean and VBAC delivery: 1989 and 1993. Stat Bull Metrop Insur Co 1996;77(1):28-36.**
**436. Gregory KD, Henry OA, Gellens AJ, et al. Repeat cesareans: how many are elective? Obstet Gynecol 1994;84(4):574-8.**

**437. Keeler EB, Park RE, Bell RM, et al. Adjusting cesarean delivery rates for case mix. Health Serv Res 1997;32(4):511-28.**
**438. Goldman G, Pineault R, Bilodeau H, et al. Effects of patient, physician and hospital characteristics on the likelihood of vaginal birth after previous cesarean section in Quebec. CMAJ 1990;143(10):1017-24.**
**439. Oleske DM, Glandon GL, Giacomelli GJ, et al. The cesarean birth rate: influence of hospital teaching status. Health Serv Res 1991;26(3):325-37.**
**440. Kazandjian VA, Lied TR. Cesarean section rates: effects of participation in a performance measurement project. Jt Comm J Qual Improv 1998;24(4):187-96.**
**441. Gregory KD, Hackmeyer P, Gold L, et al. Using the continuous quality improvement process to safely lower the cesarean section rate. Jt Comm J Qual Improv 1999;25(12):619-29.**
**442. Bickell NA, Zdeb MS, Applegate MS, et al. Effect of external peer review on cesarean delivery rates: a statewide program. Obstet Gynecol 1996;87(5 Pt 1):664-7.**
**443. Joint Commission on Accreditation of HealthCare Organizations. ,Joint Commission on Accreditation of HealthCare Organizations.**
**444. Southern Surgeons Club. A prospective analysis of 1518 laparoscopic cholecystectomies. NEJM 1991;324:1073-1078.**
**445. Karayiannakis AJ, Makri GG, Mantzioka A, et al. Postoperative pulmonary function after laparoscopic and open cholecystectomy. Br J Anaesth 1996;77(4):448-52.**
**446. Berggren U, Gordh T, Grama D, et al. Laparoscopic versus open cholecystectomy: hospitalization, sick leave, analgesia and trauma responses [see comments]. Br J Surg 1994;81(9):1362-5.**
**447. Barkun JS, Barkun AN, Sampalis JS, et al. Randomised controlled trial of laparoscopic versus mini cholecystectomy. The McGill Gallstone Treatment Group [see comments]. Lancet 1992;340(8828):1116-9.**
**448. Trondsen E, Reiertsen O, Andersen OK, et al. Laparoscopic and open cholecystectomy. A prospective, randomized study. Eur J Surg 1993;159(4):217-21.**
**449. McGinn FP, Miles AJ, Uglow M, et al. Randomized trial of laparoscopic cholecystectomy and mini- cholecystectomy [see comments]. Br J Surg 1995;82(10):1374-7.**
**450. McMahon AJ, Russell IT, Baxter JN, et al. Laparoscopic versus minilaparotomy cholecystectomy: a randomised trial [see comments]. Lancet 1994;343(8890):135-8.**
**451. McMahon AJ, Russell IT, Ramsay G, et al. Laparoscopic and minilaparotomy cholecystectomy: a randomized trial comparing postoperative pain and pulmonary function. Surgery 1994;115(5):533-9.**
**452. Steiner CA, Bass EB, Talamini MA, et al. Surgical rates and operative mortality for open and laparoscopic cholecystectomy in Maryland. N Engl J Med 1994;330(6):403-8.**
**453. Soper NJ, Barteau JA, Clayman RV, et al. Comparison of early postoperative results for laparoscopic versus standard open cholecystectomy. Surg Gynecol Obstet 1992;174(2):114-8.**
**454. NIH Consensus Development Panel on Gallstones and Laparoscopic Cholecystectomy: Gallstones and laparoscopic cholecystectomy. JAMA 1996;269:1018-1024.**
**455. Feldman MG, Russell JC, Lynch JT, et al. Comparison of mortality rates for open and closed cholecystectomy in the elderly: Connecticut statewide survey. J Laparoendosc Surg 1994;4(3):165-72.**

**456. Ferreira MR, Bennett RL, Gilman SC, et al. Diffusion of laparoscopic cholecystectomy in the Veterans Affairs health care system, 1991-1995. Eff Clin Pract 1999;2(2):49-55.**
**457. Majeed AW, Troy G, Nicholl JP, et al. Randomised, prospective, single-blind comparison of laparoscopic versus small-incision cholecystectomy [see comments]. Lancet 1996;347(9007):989-94.**
**458. Guidelines for the clinical application of laparoscopic biliary tract surgery. Surg Endosc 1998;12:191-192.**
**459. Jatzko GR, Lisborg PH, Pertl AM, et al. Multivariate comparison of complications after laparoscopic cholecystectomy and open cholecystectomy. Ann Surg 1995;221(4):381-6.**
**460. Fletcher DR, Hobbs MS, Tan P, et al. Complications of cholecystectomy: risks of the laparoscopic approach and protective effects of operative cholangiography: a population-based study. Ann Surg 1999;229(4):449-57.**
**461. Matthews BD, Williams GB. Laparoscopic cholecystectomy in an academic hospital: evaluation of changes in perioperative outcomes. Jsls 1999;3(1):9-17.**
**462. Holcomb GW, 3rd, Morgan WM, 3rd, Neblett WW, 3rd, et al. Laparoscopic cholecystectomy in children: lessons learned from the first 100 patients. J Pediatr Surg 1999;34(8):1236-40.**
**463. Maxwell JG, Tyler BA, Rutledge R, et al. Cholecystectomy in patients aged 80 and older. Am J Surg 1998;176(6):627-31.**
**464. Fendrick AM, Escarce JJ, McLane C, et al. Hospital adoption of laparoscopic cholecystectomy. Med Care 1994;32(10):1058-63.**
**465. Nenner RP, Imperato PJ, Rosenberg C, et al. Increased cholecystectomy rates among Medicare patients after the introduction of laparoscopic cholecystectomy. J Community Health 1994;19(6):409-15.**
**466. Steinle EW, VanderMolen RL, Silbergleit A, et al. Impact of laparoscopic cholecystectomy on indications for surgical treatment of gallstones. Surg Endosc 1997;11(9):933-5.**
**467. Shea JA, Berlin JA, Bachwich DR, et al. Indications for and outcomes of cholecystectomy: a comparison of the pre and postlaparoscopic eras. Ann Surg 1998;227(3):343-50.**
**468. Escarce JJ, Chen W, Schwartz JS. Falling cholecystectomy thresholds since the introduction of laparoscopic cholecystectomy . JAMA 1995;273(20):1581-5.**
**469. Marshall D, Clark E, Hailey D. The impact of laparoscopic cholecystectomy in Canada and Australia. Health Policy 1994;26(3):221-30.**
**470. Lam C, Murray F, Cuschieri A. Increased cholecystectomy rates after the introduction of laparoscopic cholecystectomy in Scotland. Gut 1996;38:282-284.**
**471. Cagir B, Rangraj M, Maffuci L, et al. The learning curve for laparoscopic cholecystectomy. J Laparoendosc Surg 1994;4(6):419-27.**
**472. Kane R, Lurie N, Borbas Cea. The outcomes of elective laparoscopic and open cholecystectomies. J Am Coll Surg 1995;180(136-145).**
**473. Orlando Rd, Russell JC, Lynch J, et al. Laparoscopic cholecystectomy. A statewide experience. The Connecticut Laparoscopic Cholecystectomy Registry. Arch Surg 1993;128(5):494-8; discussion 498-9.**

**474. Gittelsohn A, Powe NR. Small area variations in health care delivery in Maryland. Health Serv Res 1995;30(2):295-317.**
**475. Leape LL, Hilborne LH, Schwartz JS, et al. The appropriateness of coronary artery bypass graft surgery in academic medical centers. Working Group of the Appropriateness Project of the Academic Medical Center Consortium. Ann Intern Med 1996;125(1):8-18.**
**476. Leape LL, Park RE, Bashore TM, et al. Effect of variability in the interpretation of coronary angiograms on the appropriateness of use of coronary revascularization procedures. American Heart Journal 2000;139(1 Pt 1):106-13.**
**477. McGlynn EA, Naylor CD, Anderson GM, et al. Comparison of the appropriateness of coronary angiography and coronary artery bypass graft surgery between Canada and New York State. JAMA 1994;272(12):934-40.**
**478. Bernstein SJ, Brorsson B, Aberg T, et al. Appropriateness of referral of coronary angiography patients in Sweden. SECOR/SBU Project Group. Heart 1999;81(5):470-7.**
**479. Dartmouth Atlas of Health Care. ,Center for the Evaluative Clinical Sciences at Dartmouth Medical School.**
**480. Kuh D, Stirling S. Socioeconomic variation in admission for diseases of female genital system and breast in a national cohort aged 15-43. BMJ 1995;311(7009):840-3.**
**481. Broder MS, Kanouse DE, Mittman BS, et al. The appropriateness of recommendations for hysterectomy. Obstet Gynecol 2000;95(2):199-205.**
**482. Barron M, Kazandjian VA. Geographic variation in lumbar diskectomy: a protocol for evaluation. QRB Qual Rev Bull 1992;18(3):98-107.**
**483. Milamed DR, Warfield CA, Hedley-Whyte J, et al. Laminectomy and the treatment of lower-back pain in Massachusetts. Int J Technol Assess Health Care 1993;9(3):426-39.**
**484. Larequi-Lauber T, Vader JP, Burnand B, et al. Appropriateness of indications for surgery of lumbar disc hernia and spinal stenosis. Spine 1997;22(2):203-9.**
**485. Porchet F, Vader JP, Larequi-Lauber T, et al. The assessment of appropriate indications for laminectomy. J Bone Joint Surg Br 1999;81(2):234-9.**
**486. Ryan TJ, Antman EM, Brooks NH, et al. 1999 update: ACC/AHA guidelines for the management of patients with acute myocardial infarction. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Management of Acute Myocardial Infarction). J Am Coll Cardiol 1999;34(3):890-911.**
**487. Ziskind AA, Lauer MA, Bishop G, et al. Assessing the appropriateness of coronary revascularization: the University of Maryland Revascularization Appropriateness Score (RAS) and its comparison to RAND expert panel ratings and American College of Cardiology/American Heart Association guidelines with regard to assigned appropriateness rating and ability to predict outcome. Clin Cardiol 1999;22(2):67-76.**
**488. Hilborne LH, Leape LL, Bernstein SJ, et al. The appropriateness of use of percutaneous transluminal coronary angioplasty in New York State. Jama 1993;269(6):761-5.**
**489. Billings J, Zeital L, Lukomnik J, et al. Analysis of variation in hospital admission rates associated with area income in New York City. ,Unpublished Report.**
**490. Silver MP, Babitz ME, Magill MK. Ambulatory care sensitive hospitalization rates in the aged Medicare population in Utah, 1990 to 1994: a rural-urban comparison. J Rural Health 1997;13(4):285-94.**

**491. Millman M, editor. Committee on Monitoring Access to Personal Health Care Services. Washington, D.C.: National Academy Press; 1993.**
**492. Bindman A, Grumbach K, Osmond D, et al. Accuracy of preventable hospitalization rates for measureing access to care in rural communities. JGIM 1996;11[Suppl 1]:64.**
**493. Giuffrida A, Gravelle H, Roland M. Can we use ambulatory sensitive admission rates as primary care quality indicators? Evidence from England. Abstract Book/Association for Health Services Research 1998;15:104-5.**
**494. Connell FA, Blide LA, Hanken MA. Clinical correlates of small area variations in population-based admission rates for diabetes. Med Care 1984;22(10):939-49.**
**495. Hrdy SA, Hoppe PM, Bouda DW. Nebraska outpatient care quality assessment. Nebr Med J 1993;78(2):36-41.**
**496. Komaromy M, Lurie N, Osmond D, et al. Physician practice style and rates of hospitalization for chronic medical conditions. Med Care 1996;34(6):594-609.**
**497. Epstein A. The role of the medical market in preventable hospitalizations. Abstract Book/Association of Health Services Research 1998;15(316-7).**
**498. Gill JM, Mainous AG, 3rd. The role of provider continuity in preventing hospitalizations. Arch Fam Med 1998;7(4):352-7.**
**499. Falik M, Needleman J, McCall N, et al. Ambulatory care sensitive conditions: hospitalization rates by usual source of care. Abstract Book/Association for Health Services Research 1998;15:236-7.**
**500. Gill JM. Can hospitalizations be avoided by having a regular source of care? Fam Med 1997;29(3):166-71.**
**501. Shi L, Samuels ME, Pease M, et al. Patient characteristics associated with hospitalizations for ambulatory care sensitive conditions in South Carolina. Southern Medical Journal 1999;92(10):989-98.**
**502. Krakauer H, Jacoby I, Millman M, et al. Physician impact on hospital admission and on mortality rates in the Medicare population. Health Serv Res 1996;31(2):191-211.**
**503. Kaestner R, Racine A, Joyce T. Did recent expansions in Medicaid narrow socioeconomic differences in hospitalization rates of infants? Med Care 2000;38(2):195-206.**
**504. Homan R, Bryce C. Preventable hospitalization rates as a marker for access problems. Abstract Book/Association for Health Services Research 1997;14:67.**
**505. Taroni F, Repetto F, Louis DZ, et al. Variation in hospital use and avoidable patient morbidity. J Health Serv Res Policy 1997;2(4):217-22.**
**506. Weinberg AD, Minaker KL. Dehydration. Evaluation and management in older adults. Council on Scientific Affairs, American Medical Association [see comments]. Jama 1995;274(19):1552-6.**
**507. Foster DA, Talsma A, Furumoto-Dawson A, et al. Influenza vaccine effectiveness in preventing hospitalization for pneumonia in the elderly. Am J Epidemiol 1992;136(3):296-307.**
**508. Pneumococcal and influenza vaccination levels among adults aged > or = 65 years--United States, 1995 [published erratum appears in MMWR Morb Mortal Wkly Rep 19997 Oct 17;46(41):974]. MMWR Morb Mortal Wkly Rep 1997;46(39):913-9.**
**509. Marrie TJ, Lau CY, Wheeler SL, et al. A controlled trial of a critical pathway for treatment of community- acquired pneumonia. CAPITAL Study Investigators. Community-Acquired Pneumonia Intervention Trial Assessing Levofloxacin. JAMA 2000;283(6):749-55.**

**510. Braveman P, Schaaf VM, Egerter S, et al. Insurance-related differences in the risk of ruptured appendix [see comments]. N Engl J Med 1994;331(7):444-9.**
**511. Bratton SL, Haberkern CM, Waldhausen JH. Acute appendicitis risks of complications: age and Medicaid insurance. Pediatrics 2000;106(1 Pt 1):75-8.**
**512. Blumberg MS, Juhn PI. Insurance and the risk of ruptured appendix [letter; comment]. N Engl J Med 1995;332(6):395-6; discussion 397-8.**
**513. O'Toole SJ, Karamanoukian HL, Allen JE, et al. Insurance-related differences in the presentation of pediatric appendicitis. J Pediatr Surg 1996;31(8):1032-4.**
**514. Gibbons RJ, Chatterjee K, Daley J, et al. ACC/AHA/ACP-ASIM guidelines for the management of patients with chronic stable angina: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Management of Patients With Chronic Stable Angina) [published erratum appears in J Am Coll Cardiol 1999 Jul;34(1):314]. J Am Coll Cardiol 1999;33(7):2092-197.**
**515. Braunwald E, Antman EM, Beasley JW, et al. ACC/AHA guidelines for the management of patients with unstable angina and non-ST-segment elevation myocardial infarction. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on the Management of Patients With Unstable Angina). J Am Coll Cardiol 2000;36(3):970-1062.**
**516. Adams, Marano. Current estimates from the National Health Interview Survey, 1994. Vital Health Stat 1995;10(94).**
**517. HCUPnet. ,Agency for Healthcare Research and Quality,.**
**518. National Center Health Statistics. ,Centers for Disease Control.**
**519. National Center for Health Statistics. Ambulatory care vistis to physicians' offices, hospital outpatient departments, and emergency departments: United States, 1996. Vital and Health Statistics 1998;13(134).**
**520. National Heart Lung and Blood Institute/ National Asthma Education and Prevention Program. Expert Panel Report 2: Guidelines for the diagnosis and management of asthma. In: National Institutes of Health pub. no. 97-4051. Bethesda, MD; 1997.**
**521. Blais L, Ernst P, Boivin JF, et al. Inhaled corticosteroids and the prevention of readmission to hospital for asthma. Am J Respir Crit Care Med 1998;158(1):126-32.**
**522. Donahue JG, Weiss ST, Livingston JM, et al. Inhaled steroids and the risk of hospitalization for asthma. Jama 1997;277(11):887-91.**
**523. Rosenstreich DL, Eggleston P, Kattan M, et al. The role of cockroach allergy and exposure to cockroach allergen in causing morbidity among inner-city children with asthma. N Engl J Med 1997;336(19):1356-63.**
**524. Abbey DE, Petersen F, Mills PK, et al. Long-term ambient concentrations of total suspended particulates, ozone, and sulfur dioxide and respiratory symptoms in a nonsmoking population. Arch Environ Health 1993;48(1):33-46.**
**525. Greer JR, Abbey DE, Burchette RJ. Asthma related to occupational and ambient air pollutants in nonsmokers. J Occup Med 1993;35(9):909-15.**
**526. Jindal SK, Gupta D, Singh A. Indices of morbidity and control of asthma in adult patients exposed to environmental tobacco smoke. Chest 1994;106(3):746-9.**
**527. Chilmonczyk BA, Salmun LM, Megathlin KN, et al. Association between exposure to environmental tobacco smoke and exacerbations of asthma in children. N Engl J Med 1993;328(23):1665-9.**

**528. Koenig JQ. Air pollution and asthma. J Allergy Clin Immunol 1999;104(4 Pt 1):717-22.**
**529. Abbey DE, Petersen FF, Mills PK, et al. Chronic respiratory disease associated with long-term ambient concentrations of sulfates and other air pollutants. J Expo Anal Environ Epidemiol 1993;3(Suppl 1):99-115.**
**530. Koenig JQ, Larson TV, Hanley QS, et al. Pulmonary function changes in children associated with fine particulate matter. Environ Res 1993;63(1):26-38.**
**531. Walters S, Griffiths RK, Ayres JG. Temporal association between hospital admissions for asthma in Birmingham and ambient levels of sulphur dioxide and smoke. Thorax 1994;49(2):133-40.**
**532. Schwartz J, Slater D, Larson TV, et al. Particulate air pollution and hospital emergency room visits for asthma in Seattle. Am Rev Respir Dis 1993;147(4):826-31.**
**533. Cody RP, Weisel CP, Birnbaum G, et al. The effect of ozone associated with summertime photochemical smog on the frequency of asthma visits to hospital emergency departments. Environ Res 1992;58(2):184-94.**
**534. Ponka A. Asthma and low level air pollution in Helsinki. Arch Environ Health 1991;46(5):262-70.**
**535. Thurston GD, Ito K, Kinney PL, et al. A multi-year study of air pollution and respiratory hospital admissions in three New York State metropolitan areas: results for 1988 and 1989 summers. J Expo Anal Environ Epidemiol 1992;2(4):429-50.**
**536. Romieu I, Meneses F, Sienra-Monge JJ, et al. Effects of urban air pollutants on emergency visits for childhood asthma in Mexico City. Am J Epidemiol 1995;141(6):546-53.**
**537. Kesten S, Szalai J, Dzyngel B. Air quality and the frequency of emergency room visits for asthma. Ann Allergy Asthma Immunol 1995;74(3):269-73.**
**538. White MC, Etzel RA, Wilcox WD, et al. Exacerbations of childhood asthma and ozone pollution in Atlanta. Environ Res 1994;65(1):56-68.**
**539. Anderson HR, Ponce de Leon A, Bland JM, et al. Air pollution, pollens, and daily admissions for asthma in London 1987- 92. Thorax 1998;53(10):842-8.**
**540. Sheppard L, Levy D, Norris G, et al. Effects of ambient air pollution on nonelderly asthma hospital admissions in Seattle, Washington, 1987-1994 . Epidemiology 1999;10(1):23-30.**
**541. Blanc PD, Toren K. How much adult asthma can be attributed to occupational factors? Am J Med 1999;107(6):580-7.**
**542. Chan-Yeung M, Malo JL. Aetiological agents in occupational asthma. Eur Respir J 1994;7(2):346-71.**
**543. Schenker MB, Gold EB, Lopez RL, et al. Asthma mortality in California, 1960-1989. Demographic patterns and occupational associations. Am Rev Respir Dis 1993;147(6 Pt 1):1454-60.**
**544. Von Behren J, Kreutzer R, Smith D. Asthma hospitalization trends in California, 1983-1996. J Asthma 1999;36(7):575-82.**
**545. Lin S, Fitzgerald E, Hwang SA, et al. Asthma hospitalization rates and socioeconomic status in New York State (1987-1993). J Asthma 1999;36(3):239-51.**
**546. Asthma mortality and hospitalization among children and young adults-- United States, 1980-1993. MMWR Morb Mortal Wkly Rep 1996;45(17):350-3.**

**547. Ray NF, Thamer M, Fadillioglu B, et al. Race, income, urbanicity, and asthma hospitalization in California: a small area analysis. Chest 1998;113(5):1277-84.**
**548. Blixen CE, Havstad S, Tilley BC, et al. A comparison of asthma-related healthcare use between African-Americans and Caucasians belonging to a health maintenance organization (HMO). J Asthma 1999;36(2):195-204.**
**549. Claudio L, Tulton L, Doucette J, et al. Socioeconomic factors and asthma hospitalization rates in New York City. J Asthma 1999;36(4):343-50.**
**550. Gottlieb DJ, Beiser AS, O'Connor GT. Poverty, race, and medication use are correlates of asthma hospitalization rates. A small area analysis in Boston. Chest 1995;108(1):28-35.**
**551. Giuffrida A, Gravelle H, Roland M. Measuring quality of care with routine data: avoiding confusion between performance indicators and health outcomes. Bmj 1999;319(7202):94-8.**
**552. Carr W, Zeitel L, Weiss K. Variations in asthma hospitalizations and deaths in New York City. Am J Public Health 1992;82(1):59-65.**
**553. Bierman A, Steiner C, Friedman B, et al. Ambulatory care sensitive conditions: a viable indicator of access to quality primary care? [abstract]; 1999.**
**554. Halterman JS, Aligne CA, Auinger P, et al. Inadequate therapy for asthma among children in the United States. Pediatrics 2000;105(1 Pt 3):272-6.**
**555. Murray MD, Stang P, Tierney WM. Health care use by inner-city patients with asthma. J Clin Epidemiol 1997;50(2):167-74.**
**556. Lozano P, Connell FA, Koepsell TD. Use of health services by African-American children with asthma on Medicaid. Jama 1995;274(6):469-73.**
**557. Bosco LA, Gerstman BB, Tomita DK. Variations in the use of medication for the treatment of childhood asthma in the Michigan Medicaid population, 1980 to 1986. Chest 1993;104(6):1727-32.**
**558. Hartert TV, Windom HH, Peebles RS, et al. Inadequate outpatient medical therapy for patients with asthma admitted to two urban hospitals. Am J Med 1996;100(4):386-94.**
**559. Ordonez GA, Phelan PD, Olinsky A, et al. Preventable factors in hospital admissions for asthma. Arch Dis Child 1998;78(2):143-7.**
**560. Homer CJ, Szilagyi P, Rodewald L, et al. Does quality of care affect rates of hospitalization for childhood asthma? Pediatrics 1996;98(1):18-23.**
**561. Lieu TA, Quesenberry CP, Jr., Capra AM, et al. Outpatient Management Practices Associated With Reduced Risk of Pediatric Asthma Hospitalization and Emergency Department Visits. Pediatrics 1997;100(3):334-41.**
**562. Sherrill DL, Lebowitz MD, Burrows B. Epidemiology of chronic obstructive pulmonary disease. Clin Chest Med 1990;11(3):375-87.**
**563. Sullivan SD, Ramsey SD, Lee TA. The economic burden of COPD. Chest 2000;117(2 Suppl):5S-9S.**
**564. Hackner D, Tu G, Weingarten S, et al. Guidelines in pulmonary medicine: a 25-year profile. Chest 1999;116(4):1046-62.**
**565. Ferguson GT. Recommendations for the management of COPD. Chest 2000;117(2 Suppl):23S-8S.**
**566. Feinleib M, Rosenberg HM, Collins JG, et al. Trends in COPD morbidity and mortality in the United States. Am Rev Respir Dis 1989;140(3 Pt 2):S9-18.**

**567. Osborne ML, Vollmer WM, Buist AS. Periodicity of asthma, emphysema, and chronic bronchitis in a northwest health maintenance organization. Chest 1996;110(6):1458-62.**
**568. Anthonisen NR, Connett JE, Kiley JP, et al. Effects of smoking intervention and the use of an inhaled anticholinergic bronchodilator on the rate of decline of FEV1. The Lung Health Study. Jama 1994;272(19):1497-505.**
**569. Sherrill DL, Enright P, Cline M, et al. Rates of decline in lung function among subjects who restart cigarette smoking. Chest 1996;109(4):1001-5.**
**570. Prescott E, Vestbo J. Socioeconomic status and chronic obstructive pulmonary disease. Thorax 1999;54(8):737-41.**
**571. Anderson HR, Spix C, Medina S, et al. Air pollution and daily admissions for chronic obstructive pulmonary disease in 6 European cities: results from the APHEA project . Eur Respir J 1997;10(5):1064-71.**
**572. Moolgavkar SH, Luebeck EG, Anderson EL. Air pollution and hospital admissions for respiratory causes in Minneapolis-St. Paul and Birmingham . Epidemiology 1997;8(4):364-70.**
**573. Morgan G, Corbett S, Wlodarczyk J. Air pollution and hospital admissions in Sydney, Australia, 1990 to 1994 . Am J Public Health 1998;88(12):1761-6.**
**574. Schwartz J. Air pollution and hospital admissions for the elderly in Birmingham, Alabama. Am J Epidemiol 1994;139(6):589-98.**
**575. Voigt T, Bailey M, Abramson M. Air pollution in the Latrobe Valley and its impact upon respiratory morbidity. Aust N Z J Public Health 1998;22(5):556-61.**
**576. Spix C, Anderson HR, Schwartz J, et al. Short-term effects of air pollution on hospital admissions of respiratory diseases in Europe: a quantitative summary of APHEA study results. Air Pollution and Health: a European Approach. Arch Environ Health 1998;53(1):54-64.**
**577. Schwartz J. Air pollution and hospital admissions for respiratory disease. Epidemiology 1996;7(1):20-8.**
**578. Sunyer J, Saez M, Murillo C, et al. Air pollution and emergency room admissions for chronic obstructive pulmonary disease: a 5-year study. Am J Epidemiol 1993;137(7):701-5.**
**579. Sunyer J, Anto JM, Murillo C, et al. Effects of urban air pollution on emergency room admissions for chronic obstructive pulmonary disease. Am J Epidemiol 1991;134(3):277-86; discussion 287-9.**
**580. Trudeau ME, Solano-McGuire SM. Evaluating the quality of COPD care. Am J Nurs 1999;99(3):47-50.**
**581. Ramsey SD. Suboptimal medical therapy in COPD: exploring the causes and consequences. Chest 2000;117(2 Suppl):33S-7S.**
**582. Weinberger M, Oddone EZ, Henderson WG. Does increased access to primary care reduce hospital readmissions? Veterans Affairs Cooperative Study Group on Primary Care and Hospital Readmission . N Engl J Med 1996;334(22):1441-7.**
**583. Access to Health Care in America. Washington DC: National Academy Press; 1993.**
**584. Ricketts T. Ambulatory care sensitive conditions and policy development in the United States and the United Kingdom. Abstract Book/Association for Health Services Research 1998;15:108-6.**

**585. Asch SM, Sloss EM, Hogan C, et al. Measuring underuse of necessary care among elderly medicare beneficiaries using inpatient and outpatient claims [In Process Citation]. Jama 2000;284(18):2325-33.**
**586. Edep ME, Shah NB, Tateo IM, et al. Differences between primary care physicians and cardiologists in management of congestive heart failure: relation to practice guidelines. J Am Coll Cardiol 1997;30(2):518-26.**
**587. Reis SE, Holubkov R, Edmundowicz D, et al. Treatment of patients admitted to the hospital with congestive heart failure: specialty-related disparities in practice patterns and outcomes. J Am Coll Cardiol 1997;30(3):733-8.**
**588. Philbin EF, Weil HF, Erb TA, et al. Cardiology or primary care for heart failure in the community setting: process of care and clinical outcomes. Chest 1999;116(2):346-54.**
**589. Philbin EF, Andreaou C, Rocco TA, et al. Patterns of angiotensin-converting enzyme inhibitor use in congestive heart failure in two community hospitals. Am J Cardiol 1996;77(10):832-8.**
**590. UK National Health Service High Level Performance Indicators. ,United Kingdom National Health Service.**
**591. Health Care Financing Administration. Project to Develop and Evaluate Methods to Promote Ambulatory Care Quality (DEMPAQ). .**
**592. Morris AD, Boyle DI, McMahon AD, et al. Adherence to insulin treatment, glycaemic control, and ketoacidosis in insulin-dependent diabetes mellitus. The DARTS/MEMO Collaboration. Diabetes Audit and Research in Tayside Scotland. Medicines Monitoring Unit [see comments]. Lancet 1997;350(9090):1505-10.**
**593. Day JL, Metcalfe J, Johnson P. Benefits provided by an integrated education and clinical diabetes centre: a follow-up study. Diabet Med 1992;9(9):855-9.**
**594. Hypoglycemia in the Diabetes Control and Complications Trial. The Diabetes Control and Complications Trial Research Group. Diabetes 1997;46(2):271-86.**
**595. Musey VC, Lee JK, Crawford R, et al. Diabetes in urban African-Americans. I. Cessation of insulin therapy is the major precipitating cause of diabetic ketoacidosis. Diabetes Care 1995;18(4):483-9.**
**596. Umpierrez GE, Kelly JP, Navarrete JE, et al. Hyperglycemic crises in urban blacks. Arch Intern Med 1997;157(6):669-75.**
**597. Thompson CJ, Cummings F, Chalmers J, et al. Abnormal insulin treatment behaviour: a major cause of ketoacidosis in the young adult. Diabet Med 1995;12(5):429-32.**
**598. Bagg W, Sathu A, Streat S, et al. Diabetic ketoacidosis in adults at Auckland Hospital, 1988-1996. Aust N Z J Med 1998;28(5):604-8.**
**599. Harris MI. Diabetes in America: epidemiology and scope of the problem. Diabetes Care 1998;21 Suppl 3:C11-4.**
**600. Gaster B, Hirsch IB. The effects of improved glycemic control on complications in type 2 diabetes. Arch Intern Med 1998;158(2):134-40.**
**601. Vijan S, Hofer TP, Hayward RA. Estimated benefits of glycemic control in microvascular complications in type 2 diabetes [see comments]. Ann Intern Med 1997;127(9):788-95.**
**602. Klein R, Klein BE. Relation of glycemic control to diabetic complications and health outcomes. Diabetes Care 1998;21 Suppl 3:C39-43.**

**603. Klein R, Klein BE, Moss SE. Relation of glycemic control to diabetic microvascular complications in diabetes mellitus. Ann Intern Med 1996;124(1 Pt 2):90-6.**
**604. Molyneaux LM, Constantino MI, McGill M, et al. Better glycaemic control and risk reduction of diabetic complications in Type 2 diabetes: comparison with the DCCT. Diabetes Res Clin Pract 1998;42(2):77-83.**
**605. Henry RR, Genuth S. Forum One: Current recommendations about intensification of metabolic control in non-insulin-dependent diabetes mellitus. Ann Intern Med 1996;124(1 Pt 2):175-7.**
**606. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. The Diabetes Control and Complications Trial Research Group [see comments]. N Engl J Med 1993;329(14):977-86.**
**607. Retinopathy and nephropathy in patients with type 1 diabetes four years after a trial of intensive therapy. The Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group [published erratum appears in N Engl J Med 2000 May 4;342(18):1376]. N Engl J Med 2000;342(6):381-9.**
**608. Zoorob RJ, Hagen MD. Guidelines on the care of diabetic nephropathy, retinopathy and foot disease. Am Fam Physician 1997;56(8):2021-8, 2033-4.**
**609. Diabetes preventive-care practices in managed-care organizations--Rhode Island, 1995-1996. MMWR Morb Mortal Wkly Rep 1999;48(42):958-61.**
**610. Hiss RG. Barriers to care in non-insulin-dependent diabetes mellitus. The Michigan experience. Ann Intern Med 1996;124(1 Pt 2):146-8.**
**611. National Hospital Discharge Survey. ,National Center for Health Statistics.**
**612. Newton KM, Wagner EH, Ramsey SD, et al. The use of automated data to identify complications and comorbidities of diabetes: a validation study. J Clin Epidemiol 1999;52(3):199-207.**
**613. Mayfield JA, Reiber GE, Sanders LJ, et al. Preventive foot care in people with diabetes. Diabetes Care 1998;21(12):2161-77.**
**614. Centers for Disease Control and Prevention (CDC). National Diabetes Fact Sheet: National Estimates and General Information on Diabetes in the United States. Atlanta, GA: U.S. Department of Health and Human Services (HHS), CDC,; 1999.**
**615. Pecoraro RE, Reiber GE, Burgess EM. Pathways to diabetic limb amputation. Basis for prevention. Diabetes Care 1990;13(5):513-21.**
**616. Peters AL, Legorreta AP, Ossorio RC, et al. Quality of outpatient care provided to diabetic patients. A health maintenance organization experience. Diabetes Care 1996;19(6):601-6.**
**617. Jencks SF, Cuerdon T, Burwen DR, et al. Quality of medical care delivered to Medicare beneficiaries: A profile at state and national levels. Jama 2000;284(13):1670-6.**
**618. Asch SM, Sloss EM, Hogan C, et al. Measuring underuse of necessary care among elderly Medicare beneficiaries using inpatient and outpatient claims. Jama 2000;284(18):2325-33.**
**619. Weiner JP, Parente ST, Garnick DW, et al. Variation in office-based quality. A claims-based profile of care provided to Medicare patients with diabetes. Jama 1995;273(19):1503-8.**
**620. Starfield B, Powe NR, Weiner JR, et al. Costs vs quality in different types of primary care settings. Jama 1994;272(24):1903-8.**

**621. Humphrey LL, Palumbo PJ, Butters MA, et al. The contribution of non-insulin-dependent diabetes to lower-extremity amputation in the community. Arch Intern Med 1994;154(8):885-92.**
**622. Group TG. Epidemiology of lower extremity amputation in centres in Europe, North America and East Asia. The global lower extremity amputation study group. Br J Surg 2000;87(3):328-37.**
**623. Selby JV, Zhang D. Risk factors for lower extremity amputation in persons with diabetes. Diabetes Care 1995;18(4):509-16.**
**624. Lavery LA, Ashry HR, van Houtum W, et al. Variation in the incidence and proportion of diabetes-related amputations in minorities. Diabetes Care 1996;19(1):48-52.**
**625. Reiber GE, Pecoraro RE, Koepsell TD. Risk factors for amputation in patients with diabetes mellitus. A case- control study. Ann Intern Med 1992;117(2):97-105.**
**626. Malone JM, Snyder M, Anderson G, et al. Prevention of amputation by diabetic education. Am J Surg 1989;158(6):520-3; discussion 523-4.**
**627. Patout CA, Jr., Birke JA, Horswell R, et al. Effectiveness of a comprehensive diabetes lower-extremity amputation prevention program in a predominantly low-income African-American population. Diabetes Care 2000;23(9):1339-42.**
**628. Ollendorf DA, Kotsanos JG, Wishner WJ, et al. Potential economic benefits of lower-extremity amputation prevention strategies in diabetes. Diabetes Care 1998;21(8):1240-5.**
**629. Stevens-Simon C, Orleans M. Low-birthweight prevention programs: the enigma of failure. Birth 1999;26(3):184-91.**
**630. Hessol NA, Fuentes-Afflick E, Bacchetti P. Risk of low birth weight infants among black and white parents. Obstet Gynecol 1998;92(5):814-22.**
**631. O'Campo P, Xue X, Wang MC, et al. Neighborhood risk factors for low birthweight in Baltimore: a multilevel analysis. Am J Public Health 1997;87(7):1113-8.**
**632. David RJ, Collins JW, Jr. Differing birth weight among infants of U.S.-born blacks, African-born blacks, and U.S.-born whites [see comments]. N Engl J Med 1997;337(17):1209-14.**
**633. Fuentes-Afflick E, Hessol NA, Perez-Stable EJ. Maternal birthplace, ethnicity, and low birth weight in California. Arch Pediatr Adolesc Med 1998;152(11):1105-12.**
**634. Larson EH, Hart LG, Rosenblatt RA. Is non-metropolitan residence a risk factor for poor birth outcome in the U.S.? Soc Sci Med 1997;45(2):171-88.**
**635. McDuffie RS, Jr., Beck A, Bischoff K, et al. Effect of frequency of prenatal care visits on perinatal outcome among low-risk women. A randomized controlled trial [see comments]. Jama 1996;275(11):847-51.**
**636. Murray JL, Bernfield M. The differential effect of prenatal care on the incidence of low birth weight among blacks and whites in a prepaid health care plan [see comments]. N Engl J Med 1988;319(21):1385-91.**
**637. CDC. Asthma mortality and hospitalization among children and young adults--United States, 1980-1993. MMWR Morb Mortal Wkly Rep 1996;45(17):350-3.**
**638. McCormick MC, Kass B, Elixhauser A, et al. Annual report on access to and utilization of health care for children and youth in the United States--1999. Pediatrics 2000;105(1 Pt 3):219-30.**

**639. Ali S, Osberg JS. Differences in follow-up visits between African American and white Medicaid children hospitalized with asthma. J Health Care Poor Underserved 1997;8(1):83-98.**
**640. McConnochie KM, Russo MJ, McBride JT, et al. Socioeconomic variation in asthma hospitalization: excess utilization or greater need? Pediatrics 1999;103(6):e75.**
**641. Miller JE. The effects of race/ethnicity and income on early childhood asthma prevalence and health care use. Am J Public Health 2000;90(3):428-30.**
**642. Arshad SH, Hide DW. Effect of environmental factors on the development of allergic disorders in infancy. J Allergy Clin Immunol 1992;90(2):235-41.**
**643. Frischer T, Kuehr J, Meinert R, et al. Maternal smoking in early childhood: a risk factor for bronchial responsiveness to exercise in primary-school children. J Pediatr 1992;121(1):17-22.**
**644. Schmitzberger R, Rhomberg K, Buchele H, et al. Effects of air pollution on the respiratory tract of children. Pediatr Pulmonol 1993;15(2):68-74.**
**645. Gortmaker SL, Walker DK, Jacobs FH, et al. Parental smoking and the risk of childhood asthma. Am J Public Health 1982;72(6):574-9.**
**646. Henderson FW, Henry MM, Ivins SS, et al. Correlates of recurrent wheezing in school-age children. The Physicians of Raleigh Pediatric Associates. Am J Respir Crit Care Med 1995;151(6):1786-93.**
**647. Soyseth V, Kongerud J, Boe J. Postnatal maternal smoking increases the prevalence of asthma but not of bronchial hyperresponsiveness or atopy in their children. Chest 1995;107(2):389-94.**
**648. Martinez FD, Wright AL, Taussig LM, et al. Asthma and wheezing in the first six years of life. The Group Health Medical Associates. N Engl J Med 1995;332(3):133-8.**
**649. Agudo A, Bardagi S, Romero PV, et al. Exercise-induced airways narrowing and exposure to environmental tobacco smoke in schoolchildren. Am J Epidemiol 1994;140(5):409-17.**
**650. Roorda RJ, Gerritsen J, Van Aalderen WM, et al. Risk factors for the persistence of respiratory symptoms in childhood asthma. Am Rev Respir Dis 1993;148(6 Pt 1):1490-5.**
**651. Greineder DK, Loane KC, Parks P. A randomized controlled trial of a pediatric asthma outreach program. J Allergy Clin Immunol 1999;103(3 Pt 1):436-40.**
**652. McConnochie KM, Conners GP, Lu E, et al. How commonly are children hospitalized for dehydration eligible for care in alternative settings? Archives of Pediatrics and Adolescent Medicine 1999;153(12):1233-41.**
**653. Burkhart DM. Management of acute gastroenteritis in children . American Family Physician 1999;60(9):2555-63, 2565-6.**
**654. Bezerra JA, Stathos TH, Duncan B, et al. Treatment of infants with acute diarrhea: what's recommended and what's practiced. Pediatrics 1992;90(1 Pt 1):1-4.**
**655. O'Laughlin E, Notaras E, McCullogh C, et al. Home-based management of children hospitalized with acute gastroenteritis. J Paediatr Child Health 1995;31(3):189-191.**
**656. Snyder JD. Use and misuse of oral therapy for diarrhea: comparison of US practices with American Academy of Pediatrics recommendations . Pediatrics 1991;87(1):28-33.**
**657. To T, Feldman W, Young W, et al. Hospitalization rates of children with gastroenteritis in Ontario. Can J Public Health 1996;87(1):62-5.**
**658. American Heart Association. Heart and Stroke Facts:1996 Statistical Supplement. Dallas, TX: American Heart Association; 1996.**

**659. Second Report of the California Hospitals Outcomes Project, May 1996, Acute Myocardial Infarction. Sacramento, CA: Office of Statewide Health Planning and Development; 1996.**
**660. Normand S, CN M, KS F, et al. Development and validation of a claims based index for adjusting for risk of mortality: the case of acute myocardial infarction. J Clin Epidemiol 1995;48(2):229-243.**
**661. Meehan TP, Hennen J, Radford MJ, et al. Process and outcome of care for acute myocardial infarction among Medicare beneficiaries in Connecticut: a quality improvement demonstration project. Ann Intern Med 1995;122(12):928-36.**
**662. Kuulasmaa K, Tunstall-Pedoe H, Dobson A, et al. Estimation of contribution of changes in classic risk factors to trends in coronary-event rates across the WHO MONICA Project populations [see comments]. Lancet 2000;355(9205):675-87.**
**663. Indications for fibrinolytic therapy in suspected acute myocardial infarction: collaborative overview of early mortality and major morbidity results from all randomised trials of more than 1000 patients. Fibrinolytic Therapy Trialists' (FTT) Collaborative Group [published erratum appears in Lancet 1994 Mar 19;343(8899):742] [see comments]. Lancet 1994;343(8893):311-22.**
**664. Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico (GISSI). Lancet 1986;1(8478):397-402.**
**665. Wilcox RG, von der Lippe G, Olsson CG, et al. Trial of tissue plasminogen activator for mortality reduction in acute myocardial infarction. Anglo-Scandinavian Study of Early Thrombolysis (ASSET). Lancet 1988;2(8610):525-30.**
**666. Long-term effects of intravenous anistreplase in acute myocardial infarction: final report of the AIMS study. AIMS Trial Study Group. Lancet 1990;335(8687):427-31.**
**667. Roux S, Christeller S, Ludin E. Effects of aspirin on coronary reocclusion and recurrent ischemia after thrombolysis: a meta-analysis [see comments]. J Am Coll Cardiol 1992;19(3):671-7.**
**668. Gibbons RJ, Holmes DR, Reeder GS, et al. Immediate angioplasty compared with the administration of a thrombolytic agent followed by conservative treatment for myocardial infarction. The Mayo Coronary Care Unit and Catheterization Laboratory Groups [see comments]. N Engl J Med 1993;328(10):685-91.**
**669. Grines CL, Browne KF, Marco J, et al. A comparison of immediate angioplasty with thrombolytic therapy for acute myocardial infarction. The Primary Angioplasty in Myocardial Infarction Study Group [see comments]. N Engl J Med 1993;328(10):673-9.**
**670. A clinical trial comparing primary coronary angioplasty with tissue plasminogen activator for acute myocardial infarction. The Global Use of Strategies to Open Occluded Coronary Arteries in Acute Coronary Syndromes (GUSTO IIb) Angioplasty Substudy Investigators. N Engl J Med 1997;336(23):1621-8. Published erratum appears in N Engl J Med 1997;337(4):287.**
**671. ISIS-4: a randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58,050 patients with suspected acute myocardial infarction. ISIS-4 (Fourth International Study of Infarct Survival) Collaborative Group [see comments]. Lancet 1995;345(8951):669-85.**
**672. Ambrosioni E, Borghi C, Magnani B. The effect of the angiotensin-converting-enzyme inhibitor zofenopril on mortality and morbidity after anterior myocardial**

**infarction. The Survival of Myocardial Infarction Long-Term Evaluation (SMILE) Study Investigators [see comments]. N Engl J Med 1995;332(2):80-5.**
**673. Oral captopril versus placebo among 13,634 patients with suspected acute myocardial infarction: interim report from the Chinese Cardiac Study (CCS-1) [see comments]. Lancet 1995;345(8951):686-7.**
**674. O'Connor GT, Quinton HB, Traven ND, et al. Geographic variation in the treatment of acute myocardial infarction: the Cooperative Cardiovascular Project. JAMA 1999;281(7):627-33.**
**675. Berger AK, Edris DW, Breall JA, et al. Resource use and quality of care for Medicare patients with acute myocardial infarction in Maryland and the District of Columbia: analysis of data from the Cooperative Cardiovascular Project. Am Heart J 1998;135(2 Pt 1):349-56.**
**676. Kahn KL, Rogers WH, Rubenstein LV, et al. Measuring quality of care with explicit process criteria before and after implementation of the DRG-based prospective payment system. JAMA 1990;264(15):1969-73.**
**677. Allison JJ, Kiefe CI, Weissman NW, et al. Relationship of hospital teaching status with quality of care and mortality for Medicare patients with acute MI. Jama 2000;284(10):1256-62.**
**678. Holmboe ES, Meehan TP, Radford MJ, et al. Use of critical pathways to improve the care of patients with acute myocardial infarction . Am J Med 1999;107(4):324-31.**
**679. Soumerai SB, McLaughlin TJ, Gurwitz JH, et al. Effect of local medical opinion leaders on quality of care for acute myocardial infarction: a randomized controlled trial. JAMA 1998;279(17):1358-63.**
**680. California Hospital Outcomes Project Web Page. ,Office of Statewide Health Planning and Development, State of California.**
**681. Pennsylvania Health Care Cost Containment Council. ,Pennsylvania Health Care Cost Containment Council.**
**682. Smith WM. Epidemiology of congestive heart failure. Am J Cardiol 1985;55(2):3A-8A.**
**683. Yusuf S, Thom T, Abbott RD. Changes in hypertension treatment and in congestive heart failure mortality in the United States. Hypertension 1989;13(5 Suppl):I74-9.**
**684. Croft JB, Giles WH, Pollard RA, et al. Heart failure survival among older adults in the United States: a poor prognosis for an emerging epidemic in the Medicare population [see comments]. Arch Intern Med 1999;159(5):505-10.**
**685. MacIntyre K, Capewell S, Stewart S, et al. Evidence of improving prognosis in heart failure: trends in case fatality in 66 547 patients hospitalized between 1986 and 1995 [see comments]. Circulation 2000;102(10):1126-31.**
**686. Goff DC, Jr., Pandey DK, Chan FA, et al. Congestive heart failure in the United States: is there more than meets the I(CD code)? The Corpus Christi Heart Project. Arch Intern Med 2000;160(2):197-202.**
**687. Psaty BM, Boineau R, Kuller LH, et al. The potential costs of upcoding for heart failure in the United States. Am J Cardiol 1999;84(1):108-9, A9.**
**688. Polanczyk CA, Rohde LE, Dec GW, et al. Ten-year trends in hospital care for congestive heart failure: improved outcomes and increased use of resources. Arch Intern Med 2000;160(3):325-32.**

**689. Ghali JK, Cooper R, Ford E. Trends in hospitalization rates for heart failure in the United States, 1973-1986. Evidence for increasing population prevalence [see comments]. Arch Intern Med 1990;150(4):769-73.**
**690. Senni M, Tribouilloy CM, Rodeheffer RJ, et al. Congestive heart failure in the community: a study of all incident cases in Olmsted County, Minnesota, in 1991. Circulation 1998;98(21):2282-9.**
**691. Ni H, Hershberger RE. Was the decreasing trend in hospital mortality from heart failure attributable to improved hospital care? The Oregon experience, 1991-1995. Am J Manag Care 1999;5(9):1105-15.**
**692. Selker HP, Griffith JL, D'Agostino RB. A time-insensitive predictive instrument for acute hospital mortality due to congestive heart failure: development, testing, and use for comparing hospitals: a multicenter study. Med Care 1994;32(10):1040-52.**
**693. Ho KK, Anderson KM, Kannel WB, et al. Survival after the onset of congestive heart failure in Framingham Heart Study subjects [see comments]. Circulation 1993;88(1):107-15.**
**694. Croft JB, Giles WH, Pollard RA, et al. National trends in the initial hospitalization for heart failure. J Am Geriatr Soc 1997;45(3):270-5.**
**695. Polanczyk CA, Rohde LE, Philbin EA, et al. A new casemix adjustment index for hospital mortality among patients with congestive heart failure. Med Care 1998;36(10):1489-99.**
**696. Mulrow CD, Mulrow JP, Linn WD, et al. Relative efficacy of vasodilator therapy in chronic congestive heart failure. Implications of randomized trials. Jama 1988;259(23):3422-6.**
**697. Heart failure treatment with angiotensin-converting enzyme inhibitors in hospitalized Medicare patients in 10 large states. The Large State Peer Review Organization Consortium. Arch Intern Med 1997;157(10):1103-8.**
**698. Senni M, Rodeheffer RJ, Tribouilloy CM, et al. Use of echocardiography in the management of congestive heart failure in the community [see comments]. J Am Coll Cardiol 1999;33(1):164-70.**
**699. Rosenthal GE, Baker DW, Norris DG, et al. Relationships between in-hospital and 30-day standardized hospital mortality: implications for profiling hospitals. Health Serv Res 2000;34(7):1449-68.**
**700. Rockall TA, Logan RF, Devlin HB, et al. Variation in outcome after acute upper gastrointestinal haemorrhage. The National Audit of Acute Upper Gastrointestinal Haemorrhage. Lancet 1995;346(8971):346-50.**
**701. Cooper GS, Chak A, Way LE, et al. Early endoscopy in upper gastrointestinal hemorrhage: associations with recurrent bleeding, surgery, and length of hospital stay. Gastrointest Endosc 1999;49(2):145-52.**
**702. Longstreth GF. Epidemiology and outcome of patients hospitalized with acute lower gastrointestinal hemorrhage: a population-based study. Am J Gastroenterol 1997;92(3):419-24.**
**703. Rockall TA, Logan RF, Devlin HB, et al. Selection of patients for early discharge or outpatient care after acute upper gastrointestinal haemorrhage. National Audit of Acute Upper Gastrointestinal Haemorrhage. Lancet 1996;347(9009):1138-40.**
**704. Rockall TA, Logan RF, Devlin HB, et al. Incidence of and mortality from acute upper gastrointestinal haemorrhage in the United Kingdom. Steering Committee and**

**members of the National Audit of Acute Upper Gastrointestinal Haemorrhage. Bmj 1995;311(6999):222-6.**
**705. Hay JA, Lyubashevsky E, Elashoff J, et al. Upper gastrointestinal hemorrhage clinical--guideline determining the optimal hospital length of stay. Am J Med 1996;100(3):313-22.**
**706. Cook DJ, Guyatt GH, Salena BJ, et al. Endoscopic therapy for acute nonvariceal upper gastrointestinal hemorrhage: a meta-analysis. Gastroenterology 1992;102(1):139-48.**
**707. Forsen L, Sogaard AJ, Meyer HE, et al. Survival after hip fracture: short- and long-term excess mortality according to age and gender. Osteoporos Int 1999;10(1):73-8.**
**708. Wolinsky FD, Fitzgerald JF, Stump TE. The effect of hip fracture on mortality, hospitalization, and functional status: a prospective study. Am J Public Health 1997;87(3):398-403.**
**709. Magaziner J, Lydick E, Hawkes W, et al. Excess mortality attributable to hip fracture in white women aged 70 years and older. Am J Public Health 1997;87(10):1630-6.**
**710. Fisher ES, Baron JA, Malenka DJ, et al. Hip fracture incidence and mortality in New England. Epidemiology 1991;2(2):116-22.**
**711. Popovic JR, Kozak LJ. National hospital discharge survey: annual summary, 1998 [In Process Citation]. Vital Health Stat 13 2000(148):1-194.**
**712. Romano P. .October 2000.**
**713. Myers AH, Robinson EG, Van Natta ML, et al. Hip fractures among the elderly: factors associated with in-hospital mortality. Am J Epidemiol 1991;134(10):1128-37.**
**714. Todd CJ, Freeman CJ, Camilleri-Ferrante C, et al. Differences in mortality after fracture of hip: the east Anglian audit. BMJ 1995;310(6984):904-8.**
**715. Lavernia CJ. Hemiarthroplasty in hip fracture care: effects of surgical volume on short-term outcome. J Arthroplasty 1998;13(7):774-8.**
**716. Riley G, Lubitz J. Outcomes of surgery among the Medicare aged: surgical volume and mortality. Health Care Financ Rev 1985;7(1):37-47.**
**717. Boles M. A causal model of hospital volume, structure, and process indicators and surgical outcomes. [PhD Dissertation]. Richmond, VA: Virginia Commonwealth University; 1994.**
**718. Maerki SC, Luft HS, Hunt SS. Selecting categories of patients for regionalization. Implications of the relationship between volume and outcome. Med Care 1986;24(2):148-58.**
**719. Flood AB, Scott WR, Ewy W. Does practice make perfect? Part I: The relation between hospital volume and outcomes for selected diagnostic categories. Med Care 1984;22(2):98-114.**
**720. Flood AB, Scott WR, Ewy W. Does practice make perfect? Part II: The relation between volume and outcomes and other hospital characteristics. Med Care 1984;22(2):115-25.**
**721. Perez JV, Warwick DJ, Case CP, et al. Death after proximal femoral fracture--an autopsy study. Injury 1995;26(4):237-40.**
**722. Handoll HH, Farrar MJ, McBirnie J, et al. Heparin, low molecular weight heparin and physical methods for preventing deep vein thrombosis and pulmonary embolism following surgery for hip fractures. Cochrane Database Syst Rev 2000;2.**

**723. Prevention of pulmonary embolism and deep vein thrombosis with low dose aspirin: Pulmonary Embolism Prevention (PEP) trial [see comments]. Lancet 2000;355(9212):1295-302.**
**724. Nettleman MD, Alsip J, Schrader M, et al. Predictors of mortality after acute hip fracture. J Gen Intern Med 1996;11(12):765-7.**
**725. Mangano DT, Layug EL, Wallace A, et al. Effect of atenolol on mortality and cardiovascular morbidity after noncardiac surgery. Multicenter Study of Perioperative Ischemia Research Group [see comments] [published erratum appears in N Engl J Med 1997 Apr 3;336(14):1039]. N Engl J Med 1996;335(23):1713-20.**
**726. Parker MJ, Urwin SC, Handoll HH, et al. General versus spinal/epidural anaesthesia for surgery for hip fractures in adults. Cochrane Database Syst Rev 2000;2.**
**727. Urwin SC, Parker MJ, Griffiths R. General versus regional anaesthesia for hip fracture surgery: a meta- analysis of randomized trials. Br J Anaesth 2000;84(4):450-5.**
**728. Avenell A, Handoll HH. Nutritional supplementation for hip fracture aftercare in the elderly (Cochrane Review). Cochrane Database Syst Rev 2000;4.**
**729. Parker MJ, Pryor GA. The timing of surgery for proximal femoral fractures. J Bone Joint Surg [Br] 1992;74(2):203-5.**
**730. O'Hara DA, Duff A, Berlin JA, et al. The effect of anesthetic technique on postoperative outcomes in hip fracture repair [see comments]. Anesthesiology 2000;92(4):947-57.**
**731. Holt EM, Evans RA, Hindley CJ, et al. 1000 femoral neck fractures: the effect of pre-injury mobility and surgical experience on outcome. Injury 1994;25(2):91-5.**
**732. Lu-Yao GL, Keller RB, Littenberg B, et al. Outcomes after displaced fractures of the femoral neck. A meta-analysis of one hundred and six published reports. J Bone Joint Surg Am 1994;76(1):15-25.**
**733. Rogers FB, Shackford SR, Keller MS. Early fixation reduces morbidity and mortality in elderly patients with hip fractures from low-impact falls. J Trauma 1995;39(2):261-5.**
**734. Roos LL, Walld RK, Romano PS, et al. Short-term mortality after repair of hip fracture. Do Manitoba elderly do worse? Med Care 1996;34(4):310-26.**
**735. Davis FM, Woolner DF, Frampton C, et al. Prospective, multi-centre trial of mortality following general or spinal anaesthesia for hip fracture surgery in the elderly. Br J Anaesth 1987;59(9):1080-8.**
**736. Parker MJ. Internal fixation or arthroplasty for displaced subcapital fractures in the elderly? Injury 1992;23(8):521-4.**
**737. Kahn KL, Keeler EB, Sherwood MJ, et al. Comparing outcomes of care before and after implementation of the DRG- based prospective payment system. JAMA 1990;264(15):1984-8.**
**738. Kosecoff J, Kahn KL, Rogers WH, et al. Prospective payment system and impairment at discharge. The 'quicker- and-sicker' story revisited. JAMA 1990;264(15):1980-3.**
**739. Palmer RM, Saywell RM, Jr., Zollinger TW, et al. The impact of the prospective payment system on the treatment of hip fractures in the elderly [see comments]. Arch Intern Med 1989;149(10):2237-41.**

**740. Fitzgerald JF, Moore PS, Dittus RS. The care of elderly patients with hip fracture. Changes since implementation of the prospective payment system [see comments]. N Engl J Med 1988;319(21):1392-7.**
**741. Fitzgerald JF, Fagan LF, Tierney WM, et al. Changing patterns of hip fracture care before and after implementation of the prospective payment system. Jama 1987;258(2):218-21.**
**742. Gerety MB, Soderholm-Difatte V, Winograd CH. Impact of prospective payment and discharge location on the outcome of hip fracture. J Gen Intern Med 1989;4(5):388-91.**
**743. Hoyert DL, Kochanek KD, Murphy SL. Deaths: final data for 1997. Natl Vital Stat Rep 1999;47(19):1-104.**
**744. Fine MJ, Hanusa BH, Lave JR, et al. Comparison of a disease-specific and a generic severity of illness measure for patients with community-acquired pneumonia. J Gen Intern Med 1995;10(7):359-68.**
**745. Fine MJ, Smith MA, Carson CA, et al. Prognosis and outcomes of patients with community-acquired pneumonia. A meta-analysis. JAMA 1996;275(2):134-41.**
**746. Hasley PB, Albaum MN, Li YH, et al. Do pulmonary radiographic findings at presentation predict mortality in patients with community-acquired pneumonia? Arch Intern Med 1996;156(19):2206-12.**
**747. Davis RB, Iezzoni LI, Phillips RS, et al. Predicting in-hospital mortality. The importance of functional status information. Med Care 1995;33(9):906-21.**
**748. Gleason PP, Meehan TP, Fine JM, et al. Associations between initial antimicrobial therapy and medical outcomes for hospitalized elderly patients with pneumonia. Arch Intern Med 1999;159(21):2562-72.**
**749. Dowell SF. The best treatment for pneumonia: new clues, but no definitive answers. Arch Intern Med 1999;159(21):2511-2.**
**750. Stahl JE, Barza M, DesJardin J, et al. Effect of macrolides as part of initial empiric therapy on length of stay in patients hospitalized with community-acquired pneumonia. Arch Intern Med 1999;159(21):2576-80.**
**751. Meehan TP, Fine MJ, Krumholz HM, et al. Quality of care, process, and outcomes in elderly patients with pneumonia . JAMA 1997;278(23):2080-4.**
**752. Whittle J, Lin CJ, Lave JR, et al. Relationship of provider characteristics to outcomes, process, and costs of care for community-acquired pneumonia. Med Care 1998;36(7):977-87.**
**753. Niederman MS, Bass JB, Jr., Campbell GD, et al. Guidelines for the initial management of adults with community-acquired pneumonia: diagnosis, assessment of severity, and initial antimicrobial therapy. American Thoracic Society. Medical Section of the American Lung Association. Am Rev Respir Dis 1993;148(5):1418-26.**
**754. Bartlett JG, Breiman RF, Mandell LA, et al. Community-acquired pneumonia in adults: guidelines for management. The Infectious Diseases Society of America. Clin Infect Dis 1998;26(4):811-38.**
**755. Quality of care improvements for patients with pneumonia. Florida Medical Quality Assurance, Inc. Eval Health Prof 1998;21(4):514-524.**
**756. Centers for Disease Control and Prevention. Report of Final Mortality Statitistics, 1996. Volume, 47, Number 9/. ,Available via the Internet from http://www.cdc.gov/nchs/products/pubs/pubd/nvsr/47-pre/47-pre.htm#NVSR47_19.**

**757. Brown RD, Whisnant JP, Sicks JD, et al. Stroke incidence, prevalence, and survival: secular trends in Rochester, Minnesota, through 1989. Stroke 1996;27(3):373-80.**
**758. Broderick J, Brott T, Kothari R, et al. The Greater Cincinnati/Northern Kentucky Stroke Study: preliminary first-ever and total incidence rates of stroke among blacks. Stroke 1998;29(2):415-21.**
**759. Thorvaldsen P, Asplund K, Kuulasmaa K, et al. Stroke incidence, case fatality, and mortality in the WHO MONICA project. World Health Organization Monitoring Trends and Determinants in Cardiovascular Disease [published errata appear in Stroke 1995 Aug;26(8):1504 and 1995 Dec;26(12):2376]. Stroke 1995;26(3):361-7.**
**760. Leibson CL, Naessens JM, Brown RD, et al. Accuracy of hospital discharge abstracts for identifying stroke. Stroke 1994;25(12):2348-55.**
**761. Benesch C, Witter DM, Jr., Wilder AL, et al. Inaccuracy of the International Classification of Diseases (ICD-9-CM) in identifying the diagnosis of ischemic cerebrovascular disease [published erratum appears in Neurology 1998 Jan;50(1):306]. Neurology 1997;49(3):660-4.**
**762. Goldstein LB. Accuracy of ICD-9-CM coding for the identification of patients with acute ischemic stroke: effect of modifier codes. Stroke 1998;29(8):1602-4.**
**763. Rosamond WD, Folsom AR, Chambless LE, et al. Stroke incidence and survival among middle-aged adults: 9-year follow- up of the Atherosclerosis Risk in Communities (ARIC) cohort. Stroke 1999;30(4):736-43.**
**764. Lanska DJ, Kryscio R. Geographic distribution of hospitalization rates, case fatality, and mortality from stroke in the United States. Neurology 1994;44(8):1541-50.**
**765. Feinberg WM. Guidelines for the management of transient ischemic attacks. Ad Hoc Committee on Guidelines for the Management of Transient Ischemic Attacks of the Stroke Council, American Heart Association. Heart Dis Stroke 1994;3(5):275-83.**
**766. Samsa GP, Bian J, Lipscomb J, et al. Epidemiology of recurrent cerebral infarction: a medicare claims-based comparison of first and recurrent strokes on 2-year survival and cost. Stroke 1999;30(2):338-49.**
**767. Johnston KC, Connors AF, Jr., Wagner DP, et al. A predictive risk model for outcomes of ischemic stroke. Stroke 2000;31(2):448-55.**
**768. Hinchey JA, Furlan AJ, Frank JI, et al. Is in-hospital stroke mortality an accurate measure of quality of care? . Neurology 1998;50(3):619-25.**
**769. Horner RD, Sloane RJ, Kahn KL. Is use of mechanical ventilation a reasonable proxy indicator for coma among Medicare patients hospitalized for acute stroke? Health Serv Res 1998;32(6):841-59.**
**770. Wong KS. Risk factors for early death in acute ischemic stroke and intracerebral hemorrhage: A prospective hospital-based study in Asia. Asian Acute Stroke Advisory Panel. Stroke 1999;30(11):2326-30.**
**771. Kalra L, Perez I, Smithard DG, et al. Does prior use of aspirin affect outcome in ischemic stroke? Am J Med 2000;108(3):205-9.**
**772. Tissue plasminogen activator for acute ischemic stroke. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. N Engl J Med 1995;333(24):1581-7.**
**773. Mohr JP. Thrombolytic therapy for ischemic stroke: from clinical trials to clinical practice. JAMA 2000;283(9):1189-91.**

**774. Clark WM, Wissman S, Albers GW, et al. Recombinant tissue-type plasminogen activator (Alteplase) for ischemic stroke 3 to 5 hours after symptom onset. The ATLANTIS Study: a randomized controlled trial. Alteplase Thrombolysis for Acute Noninterventional Therapy in Ischemic Stroke . JAMA 1999;282(21):2019-26.**
**775. Smith MA, Doliszny KM, Shahar E, et al. Delayed hospital arrival for acute stroke: the Minnesota Stroke Survey. Ann Intern Med 1998;129(3):190-6.**
**776. Kothari R, Jauch E, Broderick J, et al. Acute stroke: delays to presentation and emergency department evaluation. Ann Emerg Med 1999;33(1):3-8.**
**777. O'Connor RE, McGraw P, Edelsohn L. Thrombolytic therapy for acute ischemic stroke: why the majority of patients remain ineligible for treatment. Ann Emerg Med 1999;33(1):9-14.**
**778. Grotta J. Should thrombolytic therapy be the first-line treatment for acute ischemic stroke? t-PA--the best current option for most patients. N Engl J Med 1997;337(18):1310-3.**
**779. Albers GW, Bates VE, Clark WM, et al. Intravenous tissue-type plasminogen activator for treatment of acute stroke: the Standard Treatment with Alteplase to Reverse Stroke (STARS) study . JAMA 2000;283(9):1145-50.**
**780. Katzan IL, Furlan AJ, Lloyd LE, et al. Use of tissue-type plasminogen activator for acute ischemic stroke: the Cleveland area experience. JAMA 2000;283(9):1151-8.**
**781. He J, Whelton PK, Vu B, et al. Aspirin and risk of hemorrhagic stroke: a meta-analysis of randomized controlled trials. JAMA 1998;280(22):1930-5.**
**782. Johnson ES, Lanes SF, Wentworth CE, 3rd, et al. A metaregression analysis of the dose-response effect of aspirin on stroke. Arch Intern Med 1999;159(11):1248-53.**
**783. Barnett HJ, Eliasziw M. Aspirin benefit remains elusive in primary stroke prevention. Arch Neurol 2000;57(3):306-8.**
**784. Hart RG, Halperin JL, McBride R, et al. Aspirin for the primary prevention of stroke and other major vascular events: meta-analysis and hypotheses. Arch Neurol 2000;57(3):326-32.**
**785. CAST: randomised placebo-controlled trial of early aspirin use in 20,000 patients with acute ischaemic stroke. CAST (Chinese Acute Stroke Trial) Collaborative Group. Lancet 1997;349(9066):1641-9.**
**786. The International Stroke Trial (IST): a randomised trial of aspirin, subcutaneous heparin, both, or neither among 19435 patients with acute ischaemic stroke. International Stroke Trial Collaborative Group. Lancet 1997;349(9065):1569-81.**
**787. Yusuf S, Anand S, Avezum A, Jr., et al. Treatment for acute myocardial infarction. Overview of randomized clinical trials. Eur Heart J 1996;17 Suppl F:16-29.**
**788. How do stroke units improve patient outcomes? A collaborative systematic review of the randomized trials. Stroke Unit Trialists Collaboration. Stroke 1997;28(11):2139-44.**
**789. Stegmayr B, Asplund K, Hulter-Asberg K, et al. Stroke units in their natural habitat: can results of randomized trials be reproduced in routine clinical practice? Riks-Stroke Collaboration. Stroke 1999;30(4):709-14.**
**790. Langhorne P, Williams BO, Gilchrist W, et al. Do stroke units save lives? Lancet 1993;342(8868):395-8.**
**791. Gage BF, Boechler M, Doggette AL, et al. Adverse outcomes and predictors of underuse of antithrombotic therapy in medicare beneficiaries with chronic atrial fibrillation. Stroke 2000;31(4):822-7.**

**792. Vanzetto G, Machecourt J, Blendea D, et al. Additive value of thallium single-photon emission computed tomography myocardial imaging for prediction of perioperative events in clinically selected high cardiac risk patients having abdominal aortic surgery. Am J Cardiol 1996;77(2):143-8.**
**793. Baron JF, Mundler O, Bertrand M, et al. Dipyridamole-thallium scintigraphy and gated radionuclide angiography to assess cardiac risk before abdominal aortic surgery [see comments]. N Engl J Med 1994;330(10):663-9.**
**794. Pilcher DB, Davis JH, Ashikaga T, et al. Treatment of abdominal aortic aneurysm in an entire state over 7 1/2 years. Am J Surg 1980;139(4):487-94.**
**795. Poldermans D, Boersma E, Bax JJ, et al. The effect of bisoprolol on perioperative mortality and myocardial infarction in high-risk patients undergoing vascular surgery. Dutch Echocardiographic Cardiac Risk Evaluation Applying Stress Echocardiography Study Group [see comments]. N Engl J Med 1999;341(24):1789-94.**
**796. Ghali WA, Ash AS, Hall RE, et al. Statewide quality improvement initiatives and mortality after cardiac surgery. JAMA 1997;277(5):379-82.**
**797. Hannan EL, Stone CC, Biddle TL, et al. Public release of cardiac surgery outcomes data in New York: what do New York state cardiologists think of it? Am Heart J 1997;134(6):1120-8.**
**798. Hannan EL, Kilburn H, Jr., O'Donnell JF, et al. Adult open heart surgery in New York State. An analysis of risk factors and hospital mortality rates. JAMA 1990;264(21):2768-74.**
**799. New Jersey Department of Health and Senior Services. ,New Jersey Department of Health and Senior Services.**
**800. Higgins TL, Estafanous FG, Loop FD, et al. Stratification of morbidity and mortality outcome by preoperative risk factors in coronary artery bypass patients. A clinical severity score. Jama 1992;267(17):2344-8.**
**801. Parsonnet V, Dean D, Bernstein AD. A method of uniform stratification of risk for evaluating the results of surgery in acquired adult heart disease. Circulation 1989;79(6 Pt 2):I3-12.**
**802. Ottino G, Bergerone S, Di Leo M, et al. Aortocoronary bypass results: a discriminant multivariate analysis of risk factors of operative mortality. J Cardiovasc Surg (Torino) 1990;31(1):20-5.**
**803. Stahle E, Bergstrom R, Holmberg L, et al. Risk factors for operative mortality and morbidity in patients undergoing coronary artery bypass surgery for stable angina pectoris. Eur Heart J 1991;12(2):162-8.**
**804. Iyer VS, Russell WJ, Leppard P, et al. Mortality and myocardial infarction after coronary artery surgery. A review of 12,003 patients. Med J Aust 1993;159(3):166-70.**
**805. Mathew JP, Parks R, Savino JS, et al. Atrial fibrillation following coronary artery bypass graft surgery: predictors, outcomes, and resource utilization. MultiCenter Study of Perioperative Ischemia Research Group. Jama 1996;276(4):300-6.**
**806. Weightman WM, Gibbs NM, Sheminant MR, et al. Drug therapy before coronary artery surgery: nitrates are independent predictors of mortality and beta-adrenergic blockers predict survival. Anesth Analg 1999;88(2):286-91.**
**807. Dziuban SW, Jr., McIlduff JB, Miller SJ, et al. How a New York cardiac surgery program uses outcomes data. Ann Thorac Surg 1994;58(6):1871-6.**

**808. Rosenthal GE, Hammar PJ, Way LE, et al. Using hospital performance data in quality improvement: the Cleveland Health Quality Choice experience. Jt Comm J Qual Improv 1998;24(7):347-60.**
**809. Burack JH, Impellizzeri P, Homel P, et al. Public reporting of surgical mortality: a survey of New York State cardiothoracic surgeons. Ann Thorac Surg 1999;68(4):1195-200; discussion 1201-2.**
**810. Peterson ED, DeLong ER, Jollis JG, et al. The effects of New York's bypass surgery provider profiling on access to care and patient outcomes in the elderly. J Am Coll Cardiol 1998;32(4):993-9.**
**811. Romano PS, Zach A, Luft HS, et al. The California Hospital Outcomes Project: using administrative data to compare hospital performance. Jt Comm J Qual Improv 1995;21(12):668-82.**
**812. Osswald BR, Blackstone EH, Tochtermann U, et al. The meaning of early mortality after CABG. Eur J Cardiothorac Surg 1999;15(4):401-7.**
**813. California CABG Mortality Reporting Project. ,Office of Statewide Health Planning and Development.**
**814. New York Department of Health. ,New York Department of Health.**
**815. Le Roux P, Elliott J, Newell D, et al. Predicting outcome in poor-grade patients with subarachnoid hemorrhage: a retrospective review of 159 aggressively managed patients. J Neurosurg 1996;85(1):39-49.**
**816. Stachniak JB, Layon AJ, Day AL, et al. Craniotomy for intracranial aneurysm and subarachnoid hemorrhage. Is course, cost, or outcome affected by age? Stroke 1996;27(2):276-81.**
**817. Solomon RA, Mayer SA, Tarmey JJ. Relationship between the volume of craniotomies for cerebral aneurysm performed at New York state hospitals and in-hospital mortality. Stroke 1996;27(1):13-7.**
**818. Taylor CL, Yuan Z, Selman WR, et al. Mortality rates, hospital length of stay, and the cost of treating subarachnoid hemorrhage in older patients: institutional and geographical differences. J Neurosurg 1997;86(4):583-8.**
**819. Lanzino G, Kassell NF, Germanson TP, et al. Age and outcome after aneurysmal subarachnoid hemorrhage: why do older patients fare worse? J Neurosurg 1996;85(3):410-8.**
**820. Whisnant JP, Sacco SE, O'Fallon WM, et al. Referral bias in aneurysmal subarachnoid hemorrhage. J Neurosurg 1993;78(5):726-32.**
**821. Kreder HJ, Williams JI, Jaglal S, et al. Are complication rates for elective primary total hip arthroplasty in Ontario related to surgeon and hospital volumes? A preliminary investigation. Can J Surg 1998;41(6):431-7.**
**822. Sharrock NE, Cazan MG, Hargett MJ, et al. Changes in mortality after total hip and knee arthroplasty over a ten- year period. Anesth Analg 1995;80(2):242-8.**
**823. Dearborn JT, Harris WH. Postoperative mortality after total hip arthroplasty. An analysis of deaths after two thousand seven hundred and thirty-six procedures. J Bone Joint Surg Am 1998;80(9):1291-4.**
**824. Whittle J, Steinberg EP, Anderson GF, et al. Mortality after elective total hip arthroplasty in elderly Americans. Age, gender, and indication for surgery predict survival. Clin Orthop 1993(295):119-26.**

**825. Seagroatt V, Tan HS, Goldacre M, et al. Elective total hip replacement: incidence, emergency readmission rate, and postoperative mortality. Bmj 1991;303(6815):1431-5.**
**826. Cleves MA, Golden WE. Assessment of HCFA's 1992 Medicare hospital information report of mortality following admission for hip arthroplasty. Health Serv Res 1996;31(1):39-48.**
**827. Wroblewski BM, Siney PD, White R. Fatal pulmonary embolism after total hip arthroplasty. Seasonal variation. Clin Orthop 1992(276):222-4.**
**828. Medicare Quality of Care Surveillance System. ,Health Care Financing Administration.**
**829. Bartmus DA, Driscoll DJ, Offord KP, et al. The modified Fontan operation for children less than 4 years old. J Am Coll Cardiol 1990;15(2):429-35.**
**830. Knott-Craig CJ, Danielson GK, Schaff HV, et al. The modified Fontan operation. An analysis of risk factors for early postoperative death or takedown in 702 consecutive patients from one institution. J Thorac Cardiovasc Surg 1995;109(6):1237-43.**
**831. Fontan F, Fernandez G, Costa F, et al. The size of the pulmonary arteries and the results of the Fontan operation. J Thorac Cardiovasc Surg 1989;98(5 Pt 1):711-9; discussion 719-24.**
**832. Cetta F, Feldt RH, O'Leary PW, et al. Improved early morbidity and mortality after Fontan operation: the Mayo Clinic experience, 1987 to 1992. J Am Coll Cardiol 1996;28(2):480-6.**
**833. Iannettoni MD, Bove EL, Mosca RS, et al. Improving results with first-stage palliation for hypoplastic left heart syndrome. J Thorac Cardiovasc Surg 1994;107(3):934-40.**
**834. Gelatt M, Hamilton RM, McCrindle BW, et al. Arrhythmia and mortality after the Mustard procedure: a 30-year single-center experience. J Am Coll Cardiol 1997;29(1):194-201.**
**835. Gentles TL, Mayer JE, Jr., Gauvreau K, et al. Fontan operation in five hundred consecutive patients: factors influencing early and late outcome. J Thorac Cardiovasc Surg 1997;114(3):376-91.**
**836. Kaulitz R, Ziemer G, Luhmer I, et al. Modified Fontan operation in functionally univentricular hearts: preoperative risk factors and intermediate results. J Thorac Cardiovasc Surg 1996;112(3):658-64.**
**837. Fontan F, Kirklin JW, Fernandez G, et al. Outcome after a "perfect" Fontan operation. Circulation 1990;81(5):1520-36.**
**838. Kern JH, Hayes CJ, Michler RE, et al. Survival and risk factor analysis for the Norwood procedure for hypoplastic left heart syndrome. Am J Cardiol 1997;80(2):170-4.**
**839. Vricella LA, Dearani JA, Gundry SR, et al. Ultra fast track in elective congenital cardiac surgery. Ann Thorac Surg 2000;69(3):865-71.**

# APPENDIX 1

# Peer Reviewers

Quality experts reviewed this report and provided their suggestions to the authors. The initial report, released May 26, 2000, was reviewed and revised. The revised report, released November 30, 2000, was also reviewed and revised by a second set of reviewers. Each group of peer reviewers is listed in this appendix.

## Appendix 1. Peer Reviewers

**List of Peer Reviewers for Initial Draft (5/26/00) and Their Affiliations:**

| Name | Affiliation |
|---|---|
| Edward E. Bartlett, Ph.D. | George Washington University<br>Washington, D.C. |
| Mark S. Blumberg, M.D. | TruRisk<br>Oakland, CA |
| H. Joanna Jiang, Ph.D. | Healthcare Association of New York State<br>Rensselaer, NY |
| Denise Love, Ph.D. | National Association of Health Data Organizations<br>Salt Lake City, Utah |
| Harold S. Luft, Ph.D. | Institute for Health Policy Studies<br>University of California, San Francisco<br>San Francisco, CA |
| Jim Pearse, M.Sc | NSW Health<br>North Sydney, Australia |
| Elaine J. Power, MPP | National Quality Forum<br>Washington, DC |
| Elizabeth M. Sloss, Ph.D. | RAND Corporation<br>Arlington, Virginia |

**List of Peer Reviewers for Revised Draft (11/30/00) and Their Affiliations:**

| | |
|---|---|
| Tim Carey, M.D., MPH | Cecil G. Sheps Center for Health Services Research<br>Chapel Hill, NC |
| Camille Clifford, B.A. | New Mexico Health Policy Commission<br>New Mexico |
| Ethan Halm, M.D., MPH | Mt. Sinai School of Medicine<br>New York, NY |
| Christine Kovner, Ph.D. | New York University<br>New York, NY |
| Jarod Loeb, Ph.D. | Joint Commission on Accreditation of Healthcare Organizations<br>Oakbrook Terrace, IL |
| Harold S. Luft, Ph.D. | Institute for Health Policy Studies<br>University of California, San Francisco<br>San Francisco, CA |
| Michael Pine, M.D. | Michael Pine and Associates, Inc.<br>Chicago, IL |
| Elaine J. Power, MPP | National Quality Forum<br>Washington, DC |
| Steven Shortell, Ph.D. | University of California at Berkeley<br>Berkeley, CA |
| Sharon L. Sprenger, RHIA, CPHQ, MPA | Joint Commission on Accreditation of Healthcare Organizations<br>Oakbrook Terrace, IL |
| Elizabeth M. Sloss, Ph.D. | RAND Corporation<br>Arlington, Virginia |

# APPENDIX 2

# Contact List for Phone Interviews

List of researchers, consultants, end users, hospital associations and other groups contacted and interviewed by phone for background information on quality measurement relevant to HCUP QI Refinement Project.

## Appendix 2. List of Individuals Interviewed for Background Information

| Consultants |
| --- |
| Mark Blumberg<br>Vice President<br>TruRisk |
| Bill Mohlenbroch<br>Vice Chairman<br>Iameter |
| Michael Pine<br>President<br>Michael Pine and Associates |
| **Federal Agency** |
| Dicta Abel-Steinberg<br>Project Officer<br>HCFA |
| Marge Keyes<br>CONQUEST Director<br>Agency for Healthcare Research and Policy |
| **Health Plans** |
| Charles Cutler<br>Chief Medical Officer<br>American Association of Health Plans |
| **Hospital Associations** |
| Michael Boyson<br>Director, Data Services and Research<br>Colorado Health and Hospital Association |
| Barbara Brooks<br>Manager, Maryland Quality Indicators Project<br>Healthcare Association of New York State |
| Joanna Jiang<br>Senior Researcher<br>Healthcare Association of New York State |
| Don Nielson<br>Vice President for Quality<br>American Hospital Association |
| Bruce Spurlock<br>Senior Vice President<br>California Healthcare Association |
| **International** |
| Martin McKee<br>London School of Hygiene & Tropical Medicine |
| **Private Accreditors** |
| Jared Loeb<br>Sharon Sprenger<br>JCAHO |
| Peggy O'Kane<br>Executive Director<br>National Committee on Quality Assurance |
| David Lansky<br>President<br>Foundation for Accountability (FACCT) |

## Appendix 2. List of Individuals Interviewed for Background Information

| Private Data Group |
| --- |
| Susan Forbes<br>President and CEO<br>Hawaii Health Information Corporation |
| Flossie Wolf<br>Director of Policy and Legislative Affairs<br>Pennsylvania Health Care Cost Containment Council |
| **Professional Organizations** |
| David Bergman<br>Professor of Pediatric Medicine<br>American Academy of Pediatrics<br>Stanford University |
| Jeff Rideout<br>Vice President of Quality<br>UCSF-Stanford Healthcare |
| **Provider Organizations** |
| David Blumenthal<br>Professor<br>Partners Health Care |
| Denise Love<br>National Association of Health Data Organizations<br>University of Utah School of Medicine |
| Jim Mortimer<br>President<br>Midwest Business Group on Health |
| Catherine Kunkel<br>National Business Coalition on Health |
| Arnie Milstein, M.D.<br>Medical Director<br>Pacific Business Group on Health<br>William M. Mercer |
| **State Data Groups** |
| Jay Buechner, Ph.D.<br>Chief, Office of Health Statistics<br>Rhode Island Dept. of Health |
| Katherine Ganz, M.D., MSH<br>Director<br>Health Policy Commission, New Mexico |
| Michael Lundberg<br>Executive Director<br>Virginia Health Information |
| **Other** |
| Diane Weber<br>Director, Quality Management<br>University Hospital Consortium |
| Shirley Girouard<br>Vice President<br>National Association for Children's Health and Research Institutions |
| Tracy Miller<br>United Hospital Fund of New York |

# APPENDIX 3

# List of Abstracted Articles

Articles listed in this appendix were abstracted during the preliminary and/or full abstraction stage of the Phase 1 literature review: “Identifying indicators”.

## Appendix 3. Abstracted Articles. Preliminary abstraction by clinical domain.

Healthcare Cost and Utilization Project Quality Indicator Literature Review - Table of Clinical Domains

| Domain | Citation | Type of Measure |
|---|---|---|
| Adverse Events | Bates, et al., 1995. Evaluation of screening criteria for adverse events in medical patients, Med Care, 33, 5, 452-62;<br><br>Dartnell, et al., 1996. Hospitalisation for adverse events related to drug therapy: incidence, avoidability and costs, Med J Aust, 164, 11, 659-62;<br><br>Garcia-Martin, et al., 1997. Proportion of hospital deaths associated with adverse events, J Clin Epidemiol, 50, 12, 1319-26; | Adverse Events |
| Ambulatory Surgery | Owings and Kozak, 1998. Ambulatory and inpatient procedures in the United States, 1996, Vital Health Stat 13, 139, 1-119;<br><br>Pokras, et al., 1997. Ambulatory and inpatient procedures in the United States, 1994, Vital Health Stat 13, 132, 1-113;<br><br>Russell, et al., 1996. Acute day hospitalization as an alternative to inpatient treatment, Can J Psychiatry, 41, 10, 629-37; | Utilization |
| Avoidable Hospitalizations | Billings, et al., 1996. Recent findings on preventable hospitalizations [see comments], Health Aff (Millwood), 15, 3, 239-49;<br><br>Blustein, et al., 1998. Preventable hospitalizations and socioeconomic status, Health Aff (Millwood), 17, 2, 177-89;<br><br>Culler, et al., 1998. Factors related to potentially preventable hospitalizations among the elderly, Med Care, 36, 6, 804-17;<br><br>Fleming, 1995. Primary care, avoidable hospitalization, and outcomes of care: a literature review and methodological approach, Med Care Res Rev, 52, 1, 88-108; | Ambulatory Care Sensitive Conditions |

| | | |
|---|---|---|
| | Josephson and Karcz, 1997. The impact of physician economic incentives on admission rates of patients with ambulatory sensitive conditions: an analysis comparing two managed care structures and indemnity insurance, Am J Manag Care, 3, 1, 49-56;<br><br>Pappas, et al., 1997. Potentially avoidable hospitalizations: inequalities in rates between US socioeconomic groups, Am J Public Health, 87, 5, 811-6;<br><br>Parchman and Culler, 1994. Primary care physicians and avoidable hospitalizations [see comments], J Fam Pract, 39, 2, 123-8;<br><br>Schreiber and Zielinski, 1997. The meaning of ambulatory care sensitive admissions: urban and rural perspectives, J Rural Health, 13, 4, 276-84; | |
| Benchmarking | Kiefe, et al., 1998. Identifying achievable benchmarks of care: concepts and methodology, Int J Qual Health Care, 10, 5, 443-7;<br><br>Lagoe, 1998. Special report: mining admissions data. Benchmark admissions identify areas for improvement, Healthc Benchmarks, 5, 6, 81-90; | Benchmarking |
| Cardiovascular | Arom, et al., 1996. Patient characteristics, safety, and benefits of same-day admission for coronary artery bypass grafting, Ann Thorac Surg, 61, 4, 1136-9; discussion 1139-40;<br><br>Burns, et al., 1997. Outcomes for older men and women with congestive heart failure, J Am Geriatr Soc, 45, 3, 276-80;<br><br>Butler, et al., 1998. Frequency of low-risk hospital admissions for heart failure, Am J Cardiol, 81, 1, 41-4;<br><br>Caputo, et al., 1997. Effect of continuous quality improvement analysis on the delivery of primary percutaneous transluminal coronary angioplasty for acute myocardial infarction, Am J Cardiol, 79, 9, 1159-64<br><br>Chin, et al., 1998. Differences among geriatricians, general internists, and cardiologists in the care of patients with heart failure: a cautionary tale of quality assessment, J Am Geriatr Soc, 46, 11, 1349-54; | Utilization |

| | | |
|---|---|---|
| Cesarean Delivery | Aron, et al., 1998. Impact of risk-adjusting cesarean delivery rates when reporting hospital performance, Jama, 279, 24, 1968-72;<br><br>Bailit, et al., 1999. Risk adjustment for interhospital comparison of primary cesarean rates, Obstet Gynecol, 93, 6, 1025-30;<br><br>Brumfield, et al., 1998. 72-hour discharge after cesarean delivery: results in a selected Medicaid population, J Matern Fetal Med, 7, 2, 72-5;<br><br>Keeler, et al., 1997. Adjusting cesarean delivery rates for case mix, Health Serv Res, 32, 4, 511-28;<br><br>Kazandjian and Lied, 1998. Cesarean section rates: effects of participation in a performance measurement project, Jt Comm J Qual Improv, 24, 4, 187-96; | Utilization |
| Complications | Iezzoni, et al., 1994. Using administrative data to screen hospitals for high complication rates, Inquiry, 31, 1, 40-55;<br><br>Iezzoni, et al., 1994. Identifying complications of care using administrative data, Med Care, 32, 7, 700-15;<br><br>Kuykendall, et al., 1995. Identifying complications and low provider adherence to normative practices using administrative data, Health Serv Res, 30, 4, 531-54;<br><br>Silber and Rosenbaum, 1997. A spurious correlation between hospital mortality and complication rates: the importance of severity adjustment, Med Care, 35, 10 Suppl, OS77-92; | Complications |
| -ectomy | Bradbury, et al., 1997. Toward a systems quality paradigm: relating health outcomes, resource expenditures, and appropriateness of **cholecystectomy** patients, Health Serv Manage Res, 10, 4, 231-44;<br><br>Chang, et al., 1998. Improvement of medical care quality after implementation of a clinical path monitoring program for transurethral **prostatectomy** patients, Eur Urol, 33, 6, 523-8; Eggleston, et al., A retrospective analysis of 6,387 **cholecystectomies**, Med Prog Technol, | Utilization |

| | | |
|---|---|---|
| | 21, 2, 85-90;<br><br>Gillum, 1995. Epidemiology of carotid **endarterectomy** and cerebral arteriography in the United States, Stroke, 26, 9, 1724-8;<br><br>Klein, et al., 1996. Maintaining quality of care and patient satisfaction with radical **prostatectomy** in the era of cost containment, Urology, 48, 2, 269-76;<br><br>Kraiss, et al., 1995. Short-stay carotid **endarterectomy** is safe and cost-effective, Am J Surg, 169, 5, 512-5;<br><br>Pitt, 1995. Laparoscopic **cholecystectomy**. The Maryland experience, Surg Endosc, 9, 11, 1224-5; | |
| Geographic Variation | Alexander, et al., 1999. Do market-level hospital and physician resources affect small area variation in hospital use?, Med Care Res Rev, 56, 1, 94-117;<br><br>Ashton, et al., 1999. Geographic variations in utilization rates in Veterans Affairs hospitals and clinics [see comments], N Engl J Med, 340, 1, 32-9;<br><br>Hendryx and Rohland, 1994. A small area analysis of psychiatric hospitalizations to general hospitals. Effects of community mental health centers [see comments], Gen Hosp Psychiatry, 16, 5, 313-8;<br><br>Morris and Munasinghe, 1994. Geographic variability in hospital admission rates for respiratory disease among the elderly in the United States, Chest, 106, 4, 1172-81; | Utilization |
| High Volume | Begg, et al., 1998. Impact of hospital volume on operative mortality for major cancer surgery [see comments], Jama, 280, 20, 1747-51;<br><br>Birkmeyer, et al., 1999. Effect of hospital volume on in-hospital mortality with pancreaticoduodenectomy, Surgery, 125, 3, 250-6;<br><br>Choti, et al., 1998. Should hepatic resections be performed at high-volume referral centers?, J Gastrointest Surg, 2, 1, 11-20; | Volume-Outcome |

| | | |
|---|---|---|
| | Gordon, et al., 1999. Complex gastrointestinal surgery: impact of provider experience on clinical and economic outcomes, J Am Coll Surg, 189, 1, 46-56;<br><br>Hamilton and Ho, 1998. Does practice make perfect? Examining the relationship between hospital surgical volume and outcomes for hip fracture patients in Quebec, Med Care, 36, 6, 892-903;<br><br>Kreder, et al., 1998. Are complication rates for elective primary total hip arthroplasty in Ontario related to surgeon and hospital volumes? A preliminary investigation, Can J Surg, 41, 6, 431-7;<br><br>Lavernia and Guzman, 1995. Relationship of surgical volume to short-term mortality, morbidity, and hospital charges in arthroplasty, J Arthroplasty, 10, 2, 133-40;<br><br>Manheim, et al., 1998. Hospital vascular surgery volume and procedure mortality rates in California, 1982-1994, J Vasc Surg, 28, 1, 45-56; discussion 56-8;<br><br>Norton, et al., 1998. The effect of hospital volume on the in-hospital complication rate in knee replacement patients [see comments], Health Serv Res, 33, 5 Pt 1, 1191-210;<br><br>Phibbs, et al., 1996. The effects of patient volume and level of care at the hospital of birth on neonatal mortality, Jama, 276, 13, 1054-9;<br><br>Phillips, et al., 1995. The association of hospital volumes of percutaneous transluminal coronary angioplasty with adverse outcomes, length of stay, and charges in California, Med Care, 33, 5, 502-14;<br><br>Phillips and Luft, 1997. The policy implications of using hospital and physician volumes as "indicators" of quality of care in a changing health care environment, Int J Qual Health Care, 9, 5, 341-8;<br><br>Solomon, et al., 1996. Relationship between the volume of craniotomies for cerebral aneurysm performed at New York state hospitals and in-hospital mortality, Stroke, 27, 1, 13-7; | |

| | | |
|---|---|---|
| | Thiemann, et al., 1999. The association between hospital volume and survival after acute myocardial infarction in elderly patients [see comments], N Engl J Med, 340, 21, 1640-8; | |
| Hip fracture | Beringer, et al., 1996. Audit of surgical delay in relationship to outcome after proximal femoral fracture, Ulster Med J, 65, 1, 32-8;<br><br>Pocock, et al., 1999. The potentia l effect on hip fracture incidence of mass screening for osteoporosis [see comments], Med J Aust, 170, 10, 486-8; | Utilization |
| Hospital Stratification | Ansari, et al., 1996. Establishing thresholds for adverse patient outcomes, Int J Qual Health Care, 8, 3, 223-30;<br><br>Boscarino, 1996. Patients' perception of quality hospital care and hospital occupancy: are there biases associated with assessing quality care based on patients' perceptions?, Int J Qual Health Care, 8, 5, 467-77; | Structure |
| ICU readmissions | Cooper, et al., 1999. Are readmissions to the intensive care unit a useful measure of hospital performance?, Med Care, 37, 4, 399-408;<br><br>Darchy, et al., 1999. Iatrogenic diseases as a reason for admission to the intensive care unit: incidence, causes, and consequences, Arch Intern Med, 159, 1, 71-8;<br><br>Dexter, et al., 1996. Surgical ICU underutilization does not significantly discourage discharge, Health Serv Manage Res, 9, 4, 238-42; | Utilization |
| Length of stay | Chen and Naylor, 1994. Variation in hospital length of stay for acute myocardial infarction in Ontario, Canada, Med Care, 32, 5, 420-35;<br><br>Edward-Chandran, et al., 1996. Reduction of length of stay in an acute care psychiatric unit [published erratum appears in Can J Psychiatry 1996 Jun;41(5):319] [see comments], Can J Psychiatry, 41, 1, 49-51;<br><br>Gross, et al., 1997. Severity adjustment for length of stay: is it always necessary?, Clin Perform Qual Health Care, 5, 4, 169-72; | Utilization |

| | | |
|---|---|---|
| | Kogan, et al., 1998. Length of stay for specialized pediatric urologic care, Arch Pediatr Adolesc Med, 152, 11, 1126-31;<br><br>Liebergall, et al., 1999. Preadmission screening of patients scheduled for hip and knee replacement: impact on length of stay, Clin Perform Qual Health Care, 7, 1, 17-22;<br><br>Philbin and Roerden, 1997. Longer hospital length of stay is not related to better clinical outcomes in congestive heart failure, Am J Manag Care, 3, 9, 1285-91;<br><br>Philbin, et al., 1997. The relationship between hospital length of stay and rate of death in heart failure, Heart Lung, 26, 3, 177-86;<br><br>Riegel, et al., 1996. Effectiveness of a program of early hospital discharge of cardiac surgery patients [see comments] [published erratum appears in J Cardiovasc Nurs 1997 Apr;11(3):1], J Cardiovasc Nurs, 11, 1, 63-75;<br><br>Silber, et al., 1999. Conditional Length of Stay, Health Serv Res, 34, 1 Pt 2, 349-63;<br>Thomas, et al., Is patient length of stay related to quality of care?, Hosp Health Serv Adm, 42, 4, 489-507; | |
| Mortality | Hinchey, et al., 1998. Is in-hospital stroke mortality an accurate measure of quality of care? [see comments], Neurology, 50, 3, 619-25;<br><br>Iezzoni, et al., 1994. Chronic conditions and risk of in-hospital death, Health Serv Res, 29, 4, 435-60;<br><br>Khuri, et al., 1997. Risk adjustment of the postoperative mortality rate for the comparative assessment of the quality of surgical care: results of the National Veterans Affairs Surgical Risk Study, J Am Coll Surg, 185, 4, 315-27;<br><br>Rosenthal, et al., 1998. Variations in standardized hospital mortality rates for six common medical diagnoses: implications for profiling hospital quality, Med Care, 36, 7, 955-64;<br><br>Thomas and Hofer, 1998. Research evidence on the validity of risk-adjusted mortality rate | In-hospital Mortality |

| | | |
|---|---|---|
| | as a measure of hospital quality of care [published erratum appears in Med Care Res Rev 1999 Mar;56(1):118], Med Care Res Rev, 55, 4, 371-404;<br><br>Thomas and Hofer, 1999. Accuracy of risk-adjusted mortality rate as a measure of hospital quality of care, Med Care, 37, 1, 83-92; | |
| Neonatal Early Discharge | Behram, et al., 1998. Implementation of early discharges after uncomplicated vaginal deliveries: maternal and infant complications [see comments], South Med J, 91, 6, 541-5;<br><br>Bragg, et al., 1997. The effect of early discharge after vaginal delivery on neonatal readmission rates, Obstet Gynecol, 89, 6, 930-3;<br><br>Britton, et al., 1994. Early discharge of the term newborn: a continued dilemma, Pediatrics, 94, 3, 291-5;<br><br>Britton, 1998. Postpartum early hospital discharge and follow-up practices in Canada and the United States, Birth, 25, 3, 161-8;<br><br>Cruz, et al., 1997. Early hospital discharge of preterm very low birth weight infants, J Perinatol, 17, 1, 29-32;<br><br>Grullon and Grimes, 1997. The safety of early postpartum discharge: a review and critique, Obstet Gynecol, 90, 5, 860-5;<br><br>Kotagal, et al., 1997. Use of hospital-based services in the first three months of life: impact of an early discharge program, J Pediatr, 130, 2, 250-6;<br><br>Lee, et al., 1995. Association between duration of neonatal hospital stay and readmission rate [see comments], J Pediatr, 127, 5, 758-66;<br><br>Liu, et al., 1997. The safety of newborn early discharge. The Washington State experience [see comments] [published erratum appears in JAMA 1997 Dec 17;278(23):2067], Jama, 278, 4, 293-8;<br><br>Margolis, 1995. A critical review of studies of newborn discharge timing, Clin Pediatr | Length of Stay |

| | | |
|---|---|---|
| | (Phila), 34, 12, 626-34;<br><br>Welsh and Ludwig-Beymer, 1998. Shortened lengths of stay: ensuring continuity of care for mothers and babies, Lippincotts Prim Care Pract, 2, 3, 284-91; | |
| Nursing | Blecke and Decker, 1997. ANA quality indicators: meaningful measurement, Mich Nurse, 70, 9, 9-10;<br><br>Canavan, 1996. ANA/C pioneers project to develop quality indicators, Am Nurse, 28, 6, 19;<br><br>Lynn and Moore, 1997. Relationship between traditional quality indicators and perceptions of care, Semin Nurse Manag, 5, 4, 187-93; | Utilization |
| Patient Satisfaction | Cleary and Edgman-Levitan, 1997. Health care quality. Incorporating consumer perspectives, Jama, 278, 19, 1608-12;<br><br>Scholte op Reimer, et al., 1996. Patients' satisfaction with care after stroke: relation with characteristics of patients and care, Qual Health Care, 5, 3, 144-50; | Satisfaction |
| Pediatric - Admissions | Bertolino and Gessner, 1999. Pediatric admissions by family physicians and pediatricians in a semirural environment: implications for residency training, J Am Board Fam Pract, 12, 2, 128-32;<br><br>Chabra, et al., 1997. Hospital use by pediatric patients: implications for change, Am J Prev Med, 13, 6 Suppl, 30-7;<br><br>Dugdale, 1996. Patterns of disease among children: a simple and versatile measure of child health in communities, J Paediatr Child Health, 32, 5, 400-4;<br><br>Gadomski, et al., 1998. Impact of a Medicaid primary care provider and preventive care on pediatric hospitalization, Pediatrics, 101, 3, E1;<br><br>Goodman, et al., 1994. Why are children hospitalized? The role of non-clinical factors in pediatric hospitalizations, Pediatrics, 93, 6 Pt 1, 896-902; | Utilization |

| | | |
|---|---|---|
| | Mangione-Smith and McGlynn, 1998. Assessing the quality of healthcare provided to children, Health Serv Res, 33, 4 Pt 2, 1059-90;<br><br>Schuster, et al., 1997. Development of a quality of care measurement system for children and adolescents. Methodological considerations and comparisons with a system for adult women [see comments], Arch Pediatr Adolesc Med, 151, 11, 1085-92;<br><br>Schwartz, et al., 1999. Administrative data for quality improvement, Pediatrics, 103, 1 Suppl E, 291-301; | |
| Pediatric - Asthma | Ali and Osberg, 1997. Differences in follow-up visits between African American and white Medicaid children hospitalized with asthma, J Health Care Poor Underserved, 8, 1, 83-98;<br><br>Goodman, et al., 1998. Trends in pediatric asthma hospitalization rates: regional and socioeconomic differences, Pediatrics, 101, 2, 208-13;<br><br>Homer, et al., 1996. Does quality of care affect rates of hospitalization for childhood asthma?, Pediatrics, 98, 1, 18-23;<br><br>McConnochie, et al., 1997. Socioeconomic variation in asthma hospitalization: excess utilization or greater need?, Pediatrics, 103, 6, e75;<br><br>Payne, et al., 1995. Variations in pediatric pneumonia and bronchitis/asthma admission rates. Is appropriateness a factor?, Arch Pediatr Adolesc Med, 149, 2, 162-9;<br><br>To, et al., 1996. A cohort study on childhood asthma admissions and readmissions, Pediatrics, 98, 2 Pt 1, 191-5; | Ambulatory Care Sensitive |
| Pediatric – Diagnostic Clusters | McConnochie, et al., 1997. Avoidable morbidity in infants. A classification based on diagnoses in administrative databases, Med Care, 35, 3, 237-54;<br><br>McConnochie, et al., 1998. Diagnostic clusters in infants as child health outcomes. Variation among socioeconomic areas in one community, Eval Health Prof, 21, 3, 332-61; | Complication |

| | | |
|---|---|---|
| Pediatric - Gastroenteritis | Ardern and Lennon, 1997. Rotavirus gastroenteritis: is vaccine prevention near at hand?, N Z Med J, 110, 1055, 407-9;<br><br>Conway and Newport, 1994. Are all hospital admissions for acute gastroenteritis necessary?, J Infect, 29, 1, 5-8;<br><br>Elliott, et al., 1996. Pre-admission management of acute gastroenteritis [see comments], J Paediatr Child Health, 32, 1, 18-21;<br><br>To, et al., 1996. Hospitalization rates of children with gastroenteritis in Ontario, Can J Public Health, 87, 1, 62-5; | Ambulatory Care Sensitive |
| Pediatrics-Tonsillectomy | Lalakea, et al., 1999. Safety of pediatric short-stay tonsillectomy, Arch Otolaryngol Head Neck Surg, 125, 7, 749-52; | Utilization |
| Perforated Ulcer | Kong, et al., 1998. Prevalence and cost of hospitalization for gastrointestinal complications related to peptic ulcers with bleeding or perforation: comparison of two national databases, Am J Manag Care, 4, 3, 399-409; | Complication |
| Pneumonia | Hand, et al., 1997. Mortality and length of stay as performance indicators for pneumonia in the elderly, J Investig Med, 45, 4, 183-90;<br><br>Meehan, et al., 1997. Quality of care, process, and outcomes in elderly patients with pneumonia [see comments], Jama, 278, 23, 2080-4;<br><br>Minogue, et al., 1998. Patients hospitalized after initial outpatient treatment for community- acquired pneumonia, Ann Emerg Med, 31, 3, 376-80;<br><br>Murphy, et al., 1999. A multihospital effort to reduce inpatient lengths of stay for pneumonia, J Nurs Care Qual, 13, 5, 11-23; | Mortality<br>Length-of-stay<br>Ambulatory Care Sensitive Condition |
| Psychiatric | Ettner and Hermann, 1998. Inpatient psychiatric treatment of elderly Medicare beneficiaries, Psychiatr Serv, 49, 9, 1173-9;<br><br>Fortney, et al., 1996. Variation among VA hospitals in length of stay for treatment of | Length-of-stay |

| | | |
|---|---|---|
| | depression, Psychiatr Serv, 47, 6, 608-13;<br><br>Johnstone and Zolese, 1999. Systematic review of the effectiveness of planned short hospital stays for mental health care, Bmj, 318, 7195, 1387-90;<br><br>Kelly, et al., 1998. Factors in delays in discharge from acute-care psychiatry, Can J Psychiatry, 43, 5, 496-501; | |
| Psychiatric – Dual Diagnosis | Ames and Tuckwell, 1994. Psychiatric disorders among elderly patients in a general hospital, Med J Aust, 160, 11, 671-5;<br><br>Appleby, et al., 1997. The impact of substance use screening on a public psychiatric inpatient population, Psychiatr Serv, 48, 10, 1311-6;<br><br>Barrett, et al., 1998. Implementing and evaluating outcome indicators of performance for mental health agencies, J Healthc Qual, 20, 3, 6-13; quiz 52;<br><br>Blixen, et al., 1997. Dual diagnosis in elders discharged from a psychiatric hospital, Int J Geriatr Psychiatry, 12, 3, 307-13;<br><br>Bradley and Zarkin, 1997. An inpatient profile of patients with a substance abuse diagnosis in Maryland, J Subst Abuse Treat, 14, 2, 155-62;<br><br>Chung, et al., 1995. Racial differences in treatment of psychiatric inpatients, Psychiatr Serv, 46, 6, 586-91;<br><br>Cohen, et al., 1994. Rates and correlates of suicide attempts in first-admission psychotic patients, Acta Psychiatr Scand, 90, 3, 167-71;<br><br>Gerke, et al., 1997. Alcohol-related diseases in general hospital patients, Alcohol Alcohol, 32, 2, 179-84;<br><br>Koenig, 1998. Depression in hospitalized older patients with congestive heart failure, Gen Hosp Psychiatry, 20, 1, 29-43; | Screening |

| | | |
|---|---|---|
| | Koenig and Kuchibhatla, 1999. Use of health services by medically ill depressed elderly patients after hospital discharge, Am J Geriatr Psychiatry, 7, 1, 48-56;<br><br>Piazza, 1996. Dual diagnosis and adolescent psychiatric inpatients, Subst Use Misuse, 31, 2, 215-23; | |
| Psychiatric - Mortality | Amaddeo, et al., 1995. Mortality among patients with psychiatric illness. A ten-year case register study in an area with a community-based system of care, Br J Psychiatry, 166, 6, 783-8;<br><br>Ganesvaran and Shah, 1997. Psychiatric in-patient suicide rates: a 21-year study, Med Sci Law, 37, 3, 202-9;<br><br>Rossler, et al., 1995. Excess mortality among elderly psychiatric in-patients with organic mental disorder, Br J Psychiatry, 167, 4, 527-32; | In-hospital Mortality |
| Readmissions | Ashton, et al., 1995. The association between the quality of inpatient care and early readmission: a meta-analysis of the evidence, Med Care, 35, 10, 1044-59;<br><br>Beggs, et al., 1996. Factors related to rehospitalization within thirty days of discharge after coronary artery bypass grafting, Best Pract Benchmarking Healthc, 1, 4, 180-6;<br><br>Camberg, et al., 1997. Discharge destination and repeat hospitalizations, Med Care, 35, 8, 756-67;<br><br>Hofer and Hayward, 1995. Can early re-admission rates accurately detect poor-quality hospitals?, Med Care, 33, 3, 234-45; | Utilization |
| Rehabilitation | Cohen, et al., 1997. The development of an outcomes management system for acute medical rehabilitation, Am J Med Qual, 12, 1, 28-32; | Rehabilitation |
| Reporting of Quality Indicators | Aiken and Sloane, 1998. Advances in hospital outcomes research, J Health Serv Res Policy, 3, 4, 249-50;<br><br>Appleby, 1998. Data briefing. Performance indicators, Health Serv J, 108, 5625, 39; | Quality Indicators |

| | | |
|---|---|---|
| | Appleby, 1998. Data briefing. Performance framework, Health Serv J, 108, 5590, 34-5;<br><br>Bradley, et al., 1998. Monitoring clinical quality in Medicaid managed care, Conn Med, 62, 4, 215-20;<br><br>Chassin, 1998. Is health care ready for Six Sigma quality?, Milbank Q, 76, 4, 565-91, 510;<br><br>Epstein, 1998. Rolling down the runway: the challenges ahead for quality report cards, Jama, 279, 21, 1691-6;<br><br>Eddy, 1998. Performance measurement: problems and solutions [see comments], Health Aff (Millwood), 17, 4, 7-25;<br><br>Iezzoni, 1997. Assessing quality using administrative data, Ann Intern Med, 127, 8 Pt 2, 666-74;<br><br>Kazandjian, et al., 1996. Do performance indicators make a difference?, Jt Comm J Qual Improv, 22, 7, 482-91; | |
| Severity Adjustment | Asenjo, et al., 1994. Relationship between severity, costs and claims of hospitalized patients using the Severity of Illness Index, Eur J Epidemiol, 10, 5, 625-32;<br><br>Daley, et al., 1997. Risk adjustment of the postoperative morbidity rate for the comparative assessment of the quality of surgical care: results of the National Veterans Affairs Surgical Risk Study, J Am Coll Surg, 185, 4, 328-40;<br><br>DesHarnais, et al., 1997. Risk-adjusted quality outcome measures: indexes for benchmarking rates of mortality, complications, and readmissions, Qual Manag Health Care, 5, 2, 80-7;<br><br>DesHarnais, et al., 1997. How to use risk-adjusted quality indicators to assess hospitals, QRC Advis, 13, 5, 1, 6-8;<br><br>Goldfield, et al., 1999. The prospective risk adjustment system, J Ambulatory Care | Risk-Adjustment |

| | | |
|---|---|---|
| | Manage, 22, 2, 41-52;<br><br>Grana, et al., 1997. Measuring the quality of inpatient health care, Qual Manag Health Care, 6, 1, 61-9;<br><br>Iezzoni, 1997. The risks of risk adjustment [see comments], Jama, 278, 19, 1600-7;<br><br>Ingber, 1998. The current state of risk adjustment technology for capitation, J Ambulatory Care Manage, 21, 4, 1-28;<br><br>Krumholz, et al., 1999. Comparing AMI mortality among hospitals in patients 65 years of age and older: evaluating methods of risk adjustment, Circulation, 99, 23, 2986-92;<br><br>Polanczyk, et al., 1998. A new casemix adjustment index for hospital mortality among patients with congestive heart failure, Med Care, 36, 10, 1489-99; | |
| Spine Surgery | Davis, 1994. Increasing rates of cervical and lumbar spine surgery in the United States, 1979-1990, Spine, 19, 10, 1117-23; discussion 1123-4; | Utilization |
| Trauma | Amatangelo, et al., 1997. Analysis of patients discharged from receiving hospitals within 24 hours of air medical transport, Air Med J, 16, 2, 44-6; discussion 47;<br><br>Ary, et al., 1996. The increasing burden of pediatric firearm injuries on the emergency department, Pediatr Emerg Care, 12, 6, 391-3;<br><br>Ashbaugh, et al., 1995. The Ohio Bicycle Injury Study, Clin Pediatr (Phila), 34, 5, 256-60; Baker, et al., Demographic factors in a population-based survey of hospitalized, work-related, ocular injury, Am J Ophthalmol, 122, 2, 213-9;<br><br>Barillo, et al., 1995. Thermal trauma resulting from motor vehicle operation or maintenance, Accid Anal Prev, 27, 6, 829-33; | Utilization |
| Urological Surgery | Chang, et al., 1999. Effects of implementation of 18 clinical pathways on costs and quality of care among patients undergoing urological surgery, J Urol, 161, 6, 1858-62; | Utilization |

| | | |
|---|---|---|
| Validation of Quality Indicators | Boscarino and Chang, 1997. Commentary: inaccurate data on the quality of care may do more harm than good--an alternative approach is required, Am J Med Qual, 12, 4, 196-200; Hofer, et al., Validating quality indicators for hospital care, Jt Comm J Qual Improv, 23, 9, 455-67;<br><br>Huff, 1997. Comprehensive reliability assessment and comparison of quality indicators and their components, J Clin Epidemiol, 50, 12, 1395-404;<br><br>Mayer-Oakes and Barnes, 1997. Developing indicators for the Medicare Quality Indicator System (MQIS): challenges and lessons learned, Jt Comm J Qual Improv, 23, 7, 381-90;<br><br>McGlynn, 1998. Choosing and evaluating clinical performance measures, Jt Comm J Qual Improv, 24, 9, 470-9;<br><br>McGlynn and Asch, 1998. Developing a clinical performance measure, Am J Prev Med, 14, 3 Suppl, 14-21;<br><br>Romano and Mark, 1994. Bias in the coding of hospital discharge data and its implications for quality assessment, Med Care, 32, 1, 81-90;<br><br>Salzer, et al., 1997. Validating quality indicators. Quality as relationship between structure, process, and outcome, Eval Rev, 21, 3, 292-309; | Quality Indicators |
| Wound Infection | Hall, et al., 1998. The time of presentation of wound infection after cardiac surgery [see comments], J Qual Clin Pract, 18, 4, 227-31;<br><br>Lapsley and Vogels, 1998. Quality and cost impacts: prevention of post-operative clean wound infections, Int J Health Care Qual Assur Inc Leadersh Health Serv, 11, 6-7, 222-31;<br><br>Majoor, et al., 1999. The extraction of quality-of-care clinical indicators from State health department administrative databases, Med J Aust, 170, 9, 420-4;<br><br>Platell and Hall, 1997. The role of wound infection as a clinical indicator after colorectal surgery, J Qual Clin Pract, 17, 4, 203-7; | Complications |

# Appendix 3. Full abstraction articles


1. Ansari MZ, Collopy BT, McDonald IG. Establishing thresholds for adverse patient outcomes. Int J Qual Health Care 1996;8:223-30.
2. Ashton CM, Del Junco DJ, Souchek J, Wray NP, Mansyur CL. The association between the quality of inpatient care and early readmission: a meta-analysis of the evidence. Med Care 1997;35:1044-59.
3. Bates DW, O'Neil AC, Petersen LA, Lee TH, Brennan TA. Evaluation of screening criteria for adverse events in medical patients. Med Care 1995;33:452-62.
4. Bradley EH, Horwitz SM, Grogan CM, Roberto M. Monitoring clinical quality in Medicaid managed care. Conn Med 1998;62:215-20.
5. Butler J, Hanumanthu S, Chomsky D, Wilson JR. Frequency of low-risk hospital admissions for heart failure. Am J Cardiol 1998;81:41-4.
6. Chang PL, Wang TM, Huang ST, Hsieh ML, Tsui KH, Lai RH. Effects of implementation of 18 clinical pathways on costs and quality of care among patients undergoing urological surgery. J Urol 1999;161:1858-62.
7. Choti MA, Bowman HM, Pitt HA, et al. Should hepatic resections be performed at high-volume referral centers? J Gastrointest Surg 1998;2:11-20.
8. Conway SP, Newport MJ. Are all hospital admissions for acute gastroenteritis necessary? J Infect 1994;29:5-8.
9. Cooper GS, Sirio CA, Rotondi AJ, Shepardson LB, Rosenthal GE. Are readmissions to the intensive care unit a useful measure of hospital performance? Med Care 1999;37:399-408.
10. Elliott EJ, Backhouse JA, Leach JW. Pre-admission management of acute gastroenteritis [see comments]. J Paediatr Child Health 1996;32:18-21.
11. Gadomski A, Jenkins P, Nichols M. Impact of a Medicaid primary care provider and preventive care on pediatric hospitalization. Pediatrics 1998;101:E1.
12. Hand R, Klemka-Walden L, Inczauskis D. Mortality and length of stay as performance indicators for pneumonia in the elderly. J Investig Med 1997;45:183-90.
13. Hofer TP, Hayward RA. Can early re-admission rates accurately detect poor-quality hospitals? Med Care 1995;33:234-45.
14. Hofer TP, Bernstein SJ, Hayward RA, DeMonner S. Validating quality indicators for hospital care. Jt Comm J Qual Improv 1997;23:455-67.
15. Homer CJ, Szilagyi P, Rodewald L, et al. Does quality of care affect rates of hospitalization for childhood asthma? Pediatrics 1996;98:18-23.
16. Iezzoni LI, Shwartz M, Ash AS, Mackiernan YD. Using severity measures to predict the likelihood of death for pneumonia inpatients. J Gen Intern Med 1996;11:23-31.
17. Khuri SF, Daley J, Henderson W, et al. Risk adjustment of the postoperative mortality rate for the comparative assessment of the quality of surgical care: results of the National Veterans Affairs Surgical Risk Study. J Am Coll Surg 1997;185:315-27.
18. Klein EA, Grass JA, Calabrese DA, Kay RA, Sargeant W, O'Hara JF. Maintaining quality of care and patient satisfaction with radical prostatectomy in the era of cost containment. Urology 1996;48:269-76.
19. Majoor JW, Ibrahim JE, Cicuttini FM, Boyce NW, McNeil JJ. The extraction of quality-of-care clinical indicators from State health department administrative databases. Med J Aust 1999;170:420-4.
20. McConnochie KM, Roghmann KJ, Liptak GS. Socioeconomic variation in discretionary and mandatory hospitalization of infants: an ecologic analysis [see comments]. Pediatrics 1997;99:774-84.
21. Payne SM, Donahue C, Rappo P, et al. Variations in pediatric pneumonia and bronchitis/asthma admission rates. Is appropriateness a factor? Arch Pediatr Adolesc Med 1995;149:162-9.
22. Platell C, Hall JC. The role of wound infection as a clinical indicator after colorectal surgery. J Qual Clin Pract 1997;17:203-7.
23. Rosenthal GE, Shah A, Way LE, Harper DL. Variations in standardized hospital mortality rates for six common medical diagnoses: implications for profiling hospital quality. Med Care 1998;36:955-64.

24. Schuster MA, Asch SM, McGlynn EA, Kerr EA, Hardy AM, Gifford DS. Development of a quality of care measurement system for children and adolescents. Methodological considerations and comparisons with a system for adult women [see comments]. Arch Pediatr Adolesc Med 1997;151:1085-92.
25. Thomas JW, Hofer TP. Accuracy of risk-adjusted mortality rate as a measure of hospital quality of care. Med Care 1999;37:83-92.
26. To T, Dick P, Feldman W, Hernandez R. A cohort study on childhood asthma admissions and readmissions. Pediatrics 1996;98:191-5.
27. To T, Feldman W, Young W, Maloney SL. Hospitalization rates of children with gastroenteritis in Ontario. Can J Public Health 1996;87:62-5.

# APPENDIX 4

# Data Collection Instruments

These data collection instruments were used in the preliminary and full abstraction phases of the Phase1 literature review: “Identifying indicators”. Three forms are provided in the following order: 1.) Full article abstraction, 2.) Risk adjustment, and 3.) Initial screening. They provide a structured mechanism for data collection. However, currently most literature available does not provide the level of detail required by these forms. These forms may be useful for future projects once the health care quality literature expands to include more validation studies and detailed reports of potential indicators.

Abstractor's Initials:
Date Abstracted:

**Indicator Data Collection Instrument**
UCSF-Stanford EPC/AHCPR HCUP QI

| Measure Header | | | |
|---|---|---|---|
| Measure Set Name (e.g., HEDIS, DEMPAQ): | | | |
| Developer Subset Name | | | |
| Performance Measure Name: | | | |
| Measure Code: | *System assigned* | | |
| Type of Measure:<br>(structure/process/proxy -outcome/outcome) | | | |
| Year Published: | | | |
| Year Most Recently Updated:<br>("Update"= change in construction of measure) | | | |
| Developer<br>Organization (ie. JCAHO, NCQA)<br>Last Name, First Name<br>Affiliation (University, Dept.)<br>Address<br>Phone<br>Email | | | |
| Contact Person (<u>if different from Developer</u>)<br>Last Name, First Name<br>Affiliation (University, Dept.)<br>Address<br>Phone<br>Email | | | |
| Technical Support Available: | ❑ YES | ❑ NO | ❑ Unclear (*For published work, contact author may have information on support by commercial organization).* |
| Measure Detail | | | |
| Type of Measure:<br>The activity or area of major concentration | ❑ Mortality<br>❑ Readmission<br>❑ Complication<br>❑ length of stay<br>❑ avoidable hospitalization<br>❑ potentially overused procedure<br>❑ potentially underused procedure<br>❑ other | | |

| | | | | | |
|---|---|---|---|---|---|
| Enhancement Area: These are the areas AHCPR has asked us to focus on. You may choose more than one - as many as are applicable. | ❑ Chronic medical condition | ❑ Avoid. Hospit. | ❑ Pediatrics | ❑ New Technol. | ❑ None |
| Clinical Domain: One of these options should be chosen based on article. Check all that apply. | ❑ Medical | ❑ surgical | ❑ pediatric | ❑ obstetric | ❑ psychiatric |
| Level of Intervention | | | ❑ primary prevention (health promotion, disease prevention)<br>❑ secondary prevention (screening, early detection)<br>❑ tertiary prevention (optimal treatment of established disease to prevent complications/death) | | |
| Measure Rationale | | | . | | |
| Scoring method<br>(e.g., percentage, proportion, rank, mean, median, etc.) | | | ❑ rate (events/person-time at risk)<br>❑ ratio (events/events, where the numerator is not a subset of the denominator)<br>❑ mean<br>❑ median<br>❑ proportion/percentage (death/complication "rate")<br>❑ index or score<br>❑ rank or percentile<br>❑ range<br>❑ standard deviation or variance<br>❑ odds ratio or relative risk<br>❑ hazard rate<br>❑ z score (standardized difference)<br>❑ other | | |
| Quality Standard:<br>(if suggested *a priori* by developer or sponsor) | | | ❑ accepted benchmark<br>❑ external comparison<br>❑ institutional comparison | | |
| Numerator Statement:<br>A statement that depicts the portion of the denominator population that satisfies the condition of the performance measure to be an indicator event.<br>(ICD-9 diagnosis or procedure codes, DRG codes, MDC codes, or Revenue Center Codes) | | | | | |

| | |
|---|---|
| Denominator Statement:<br>A statement that depicts the population evaluated..<br>(ICD-9 diagnosis or procedure codes, DRG codes, MDC codes, CPT or Revenue Center Codes) | |
| Population Inclusions (numerator):<br>Additional information describing the population(s) not contained in the numerator statement. | |
| Population Inclusions (denominator):<br>Additional information describing the population(s) not contained in the denominator statement | |
| Population Exclusions (numerator):<br>Additional information describing the population(s) that should not be included in the numerator | |
| Population Exclusions (denominator):<br>Additional information describing the population(s) that should not be included in the denominator | |
| Longitudinal Data Required:<br>The measure requires the ability to match patient discharges over time | ❑ yes<br>❑ no |
| Additional Data Sources Required:<br>(e.g., area population counts, etc.)<br>Check all that apply, indicate specific data source (e.g. population/ census birth certificate ) | ❑ population/ census: _____<br>❑ chart data: _____<br>❑ lab data: _____<br>❑ pharmacy data: _____<br>❑ outpatient data: _____<br>❑ other: _____ |
| Extent of prior use:<br>(e.g., number/type of organizations) | ❑ None<br>❑ Single Site<br>❑ Single organization/ Multiple Sites<br>❑ Multiple organizations |
| Current Status | ❑ measure defined but not pilot tested<br>❑ pilot testing complete but not implemented<br>❑ implemented without pilot testing<br>❑ tested and implemented and still in use<br>❑ tested and implemented but discontinued |
| Scientific support for measure<br>(e.g., expert panel, published guideline, clinical trials, other empiric research)<br>Check all that apply | ❑ published guideline(s)<br>❑ clinician/expert panel(s)<br>❑ review of published literature (especially RCTs or equivalent)<br>❑ application or revision of pre-existing instruments or measures<br>❑ consensus within user group<br>❑ theory/concept only |

| | |
|---|---|
| Reference Citations: | *List the publications according to the style from the software program EndNote.* |
| Validity | ❑ face validity<br>❑ consensual validity<br>❑ criterion validity (gold standard measure)<br>❑ predictive validity<br>❑ construct validity (correlated with other measures in the absence of a gold standard)<br>❑ other<br>❑ none |
| Risk Adjustment: | ❑ none<br>❑ stratified/subgroup analysis<br>❑ paired/matched data at patient level<br>❑ risk-adjustment using publicly or commercially available software<br>❑ risk-adjustment devised specifically for this measure and condition |
| Abstractor Comments: | |

**Risk-Adjustment Method Data Collection Instrument**
UCSF-Stanford EPC/AHCPR HCUP QI

| Risk-Adjustment Method Header | | | |
|---|---|---|---|
| Method Name: | | | |
| Method Code: | *System assigned* | | |
| Year Developed: | | | |
| Year Most Recently Update<br>("Update"= Change in risk adjustment strategy)<br>(if applicable) | | | |
| Developer<br>Organization (ie. JCAHO, NCQA)<br>First Author<br>Affiliation (University, Dept.)<br>Address<br>Phone<br>Email | | | |
| Contact<br>Name<br>Affiliation (University, Dept.)<br>Address<br>Phone<br>Email | | | |
| Technical Support Available: | ❑ YES | ❑ NO | ❑ Unclear (inquire with developer) |
| ❑ Public Domain | | ❑ Proprietary | |

| Risk-Adjustment Method Detail | |
|---|---|
| Adjustment Rationale:<br>An explanation of why the adjustment is necessary to reduce or remove the influences of confounding patient factors | |
| Classification or analytic approach<br>(e.g., stratification (number of strata), logistic regression, linear regression, etc.)<br>Check all that apply | ❑ Stratification<br>❑ Logistic regression<br>❑ Linear regression<br>❑ Other |
| System development method<br>(e.g., empirical model, score based on empirical model, clinical judgment, etc.)<br>Check all that apply | ❑ Logistic regression<br>❑ Score based on empirical model<br>❑ A priori/ Clinical Judgement<br>❑ Other |
| Published performance: discrimination<br>(e.g., the extent to which the model predicts higher probabilities of an event for those who experience the event) | |
| Published performance: calibration<br>(e.g., does the mean of the predicted equal the mean of the actual for the entire population and for population subgroups) | |
| Co-morbidities:<br>Pre-existing diseases or conditions | |
| Severity of Illness Classification:<br>(AJCC staging, ASA-PS classification for surgical patients) | |
| Patient Demographics<br>(e.g., age and gender) | |
| Longitudinal Data Required:<br>The measure requires the ability to match patient discharges over time | ❑ Yes<br>❑ No |
| Additional Data Sources Required:<br>(e.g.,chart review, registry data)<br>Check all that apply, indicate specific data source (e.g. population/ census birth certificate ) | ❑ population/ census: _____<br>❑ chart data: _____<br>❑ lab data: _____<br>❑ pharmacy data: _____<br>❑ outpatient data: _____<br>❑ other: _____ |
| Extent of prior or current use:<br>(e.g., number/type of organizations) | ❑ None<br>❑ Single Site<br>❑ Single organization/ Multiple Sites<br>❑ Multiple organizations |

| | |
|---|---|
| Reference Citations: | *List the publications according to the style from the software program EndNote.* |
| Abstractor Comments: | |

Screener's Initials:
Date Screened:

**Indicator Data Screening Instrument**

UCSF-Stanford EPC/AHCPR HCUP QI

| Measure Header | | | | | |
|---|---|---|---|---|---|
| Lead Author: | | | | | |
| Article Title: | | | | | |
| Year Published: | | | | | |
| **Measure Detail** | | | | | |
| Type of Measure:<br>The activity or area of major concentration | ❑ Mortality<br>❑ Readmission<br>❑ Complication<br>❑ length of stay<br>❑ avoidable hospitalization<br>❑ potentially overused procedure<br>❑ potentially underused procedure<br>❑ other | | | | |
| Type of Measure:<br>(structure/process/proxy -outcome/outcome) | | | | | |
| Enhancement Area: These are the areas AHCPR has asked us to focus on. You may choose more than one - as many as are applicable. | ❑ Chronic medical condition | ❑ Avoid. Hospit. | ❑ Pediatrics | ❑ New Technol. | ❑ None |
| Clinical Domain: One of these options should be chosen based on article. | ❑ Medical | ❑ surgical | ❑ pediatric | ❑ obstetric | ❑ psychiatric |
| Measure Rationale | | | | | |
| Longitudinal Data Required:<br>The measure requires the ability to match patient discharges over time | ❑ yes<br>❑ no | | | | |
| Additional Data Sources Required:<br>(e.g., area population counts, etc.) | ❑ population/ census<br>❑ chart data<br>❑ lab data<br>❑ pharmacy data<br>❑ outpatient data<br>❑ other | | | | |

# APPENDIX 5

# Inventory of Indicators

This appendix includes a compilation of the indicators that we located during the Phase 1 literature review (Identifying indicators), and through interviews.

The indicators are organized by measure type (structural indicators; process indicators – utilization, length of stay; outcome indicators – ACSC/avoidable hospitalizations, complications, mortality) and by clinical domain (medical, surgical, chronic, obstetric, neonatal, pediatric, psychiatric). In some cases multiple definitions appear in the literature. Alternate definitions are noted.

The current users or developers include all the groups/individuals that we identified during our Phase 1 literature review and through phone interviews.

The literature and empirical columns refer to whether or not a detailed literature or empirical review was completed for the indicator. These reviews appear in the main text.

The reason for selection or exclusion refers to our selection process for determining which indicators to review extensively with the structured evaluation framework. The criteria and process are described in the methods section.

The indicator number is the number that has been assigned to each indicator as a unique identifier used throughout the report.

## Appendix 5. Inventory of Indicators

| Measure Type and Clinical Domain | Indicator Name | Current Users or Developers | Evidence Review* | | Reason for selection or exclusion | Indi-cator # in text body |
|---|---|---|---|---|---|---|
| | | | Empiri-cal | Liter-ature | | |
| Structure measures: | | | | | | |
| Volume Outcome measures: | | | | | | |
| Examples: | | | | | | |
| | Acute myocardial infarction (AMI) | Suggested uses in literature (Halm, Dudley) | | | Providers have limited control over frequency of AMI admissions, controversy over actual volume-outcome relationship | - |
| | Amputation of lower limb | | | | No recent studies | - |
| | Abdominal Aortic Aneurysm (AAA) repair | | x | x | Relatively frequent procedure, with strong established volume-outcome relationship | 1 |
| | Carotid endarectomy | | x | x | Frequent, high complication, mortality rates | 2 |
| | Cerebral aneurysm repair | | | | Small number of recent studies. | - |
| | Cholecystectomy | | | | No recent studies. Conflicting evidence as to the volume-outcome relationship. | - |
| | Coronary Artery Bypass Graft (CABG) surgery | | x | x | Frequent, PTCA alternative, documented relationship, current HCUP QI utilization | 3 |
| | Esophageal resection | | x | x | Though infrequent, the volume outcome relationship is particularly strong. | 4 |
| | Gastric surgery | | | | No recent studies. Conflicting evidence as to the volume-outcome relationship. | - |
| | Heart transplantation | | | | Small number of recent studies. | - |
| | Hepatic resection | | | | Infrequent procedure, low provider variation | - |

*An "x" placed in the empirical or literature evidence review column denotes that the review was conducted; a "p" denotes that preliminary analyses were conducted, but that the indicators' performance was poor on statistical tests of precision.

| Measure Type and Clinical Domain | Indicator Name | Current Users or Developers | Evidence Review* | | Reason for selection or exclusion | Indi-cator # in text body |
|---|---|---|---|---|---|---|
| | | | Empiri-cal | Liter-ature | | |
| | Hernia repair | | | | Conflicting evidence as to the volume-outcome relationship. | - |
| | Hip fracture | | | | Conflicting evidence as to the volume-outcome relationship. | - |
| | Human Immunodeficiency Virus (HIV) | | | | Literature showing relationship is based on data over a decade old. Recent advances and changes in treatment call into question the validity of those results in the year 2000. | - |
| | Lower extremity arterial bypass surgery | | | | Small number of recent studies, 2 out of 3 showed no volume effect. | - |
| | Pancreatic resection | | x | x | Though infrequent, volume relationship to outcome is very strong. | 5 |
| | Pediatric heart surgery | | x | x | Pediatrics is a focus area for new HCUP indicators, strong volume-outcome relationship | 6 |
| | Prostatectomy | | | | Conflicting evidence as to the volume-outcome relationship. | - |
| | Percutaneous Transluminal Coronary Angioplasty (PTCA) | | x | x | Frequent, CABG alternative, documented relationship with volume | 7 |
| | Total knee replacement | | | | Few recent studies. | - |
| | Total hip replacement | | | | Conflicting evidence as to the volume-outcome relationship. | - |

Process Measures, provider level:

Potentially overused:

Cesarean sections:

| | | | | | |
|---|---|---|---|---|---|
| Primary cesarean section | • National Center for Healthcare Statistics | | | Combined in all cesarean section | |
| Repeat cesarean section | • University Hospital Consortium | | | | |

*An “x” placed in the empirical or literature evidence review column denotes that the review was conducted; a “p” denotes that preliminary analyses were conducted, but that the indicators’ performance was poor on statistical tests of precision.

| Measure Type and Clinical Domain | Indicator Name | Current Users or Developers | Evidence Review* | | Reason for selection or exclusion | Indi-cator # in text body |
|---|---|---|---|---|---|---|
| | | | Empiri-cal | Liter-ature | | |
| | All cesarean section | • Cleveland Health Quality Choice<br>• Colorado Health and Hospital, proposed<br>• Florida Agency for Health Care Administration<br>• Greater New York Hospital Association<br>• HCUP<br>• HealthGrades.com<br>• IMSystem<br>• Maryland Quality Indicator Project<br>• Michigan Hospital Association<br>• Pacific Business Group on Health<br>• United Health Care<br>• University Hospital Consortium<br>• Virginia Health Information<br>• Washington State Community Health Information Partnership<br>• Submitted as ORYX measure | x | x | Current HCUP indicator | 8 |
| Surgical: | | | | | | |
| | Incidental appendectomy | • HCUP | x | x | Current HCUP | 9 |
| | Bilateral cardiac catheterization | • HCFA | x | x | Common procedures with clear indications and overuse documentation. | 10 |

*An “x” placed in the empirical or literature evidence review column denotes that the review was conducted; a “p” denotes that preliminary analyses were conducted, but that the indicators’ performance was poor on statistical tests of precision.

| Measure Type and Clinical Domain | Indicator Name | Current Users or Developers | Evidence Review* | | Reason for selection or exclusion | Indi-cator # in text body |
|---|---|---|---|---|---|---|
| | | | Empiri-cal | Liter-ature | | |
| | Radical mastectomy | • Literature, NIH consensus development conference | p | | Many mastectomy or lumpectomy procedures are performed on an outpatient basis. | - |
| Potentially underused: | | | | | | |
| Vaginal delivery after cesarean section (VBAC): | | | | | | |
| | Vaginal Birth After C-section (VBAC) | • Cleveland Health Quality Choice<br>• HealthGrades.com<br>• HCUP<br>• IMSystem<br>• JCAHO core measure<br>• Maryland QI Project<br>• Michigan Hospital Association<br>• University Hospital Consortium<br>• Submitted as ORYX measure | x | x | VBAC is a current HCUP measure | 11 |
| | Laparoscopic cholecystectomy | • HCUP<br>• Submitted as ORYX measure | x | x | Current HCUP with extensive literature | 12 |
| Process Measures, area level: | | | | | | |
| Potentially overused: | | | | | | |
| Surgical: | | | | | | |

*An "x" placed in the empirical or literature evidence review column denotes that the review was conducted; a "p" denotes that preliminary analyses were conducted, but that the indicators' performance was poor on statistical tests of precision.

| Measure Type and Clinical Domain | Indicator Name | Current Users or Developers | Evidence Review* | | Reason for selection or exclusion | Indi-cator # in text body |
|---|---|---|---|---|---|---|
| | | | Empiri-cal | Liter-ature | | |
| | Carotid endarterectomy | • Literature (eg., Chassin, Leape, Wong) | p | | Common procedure, with evidence of potential overuse. Low precision using HCUP NIS data. | - |
| | Cataract surgery | • Literature (Tobacman) | | | Potentially outpatient procedure | - |
| | Cholecystectomy | • Literature (Pilpel) | | | Potentially outpatient procedure | - |
| | Colonoscopy | • Literature (Froehlich) | | | Outpatient procedure | - |
| | Coronary angiography (includes diagnostic testing for coronary artery disease) | • Literature (eg., Bernstein, Carlisle Chassin, Gray, Leape) | | | Outpatient procedure | - |
| | Coronary artery bypass graft (CABG) | • HCUP<br>• Submitted as ORYX measure<br>• Literature (Gray, Leape, McGlynn) | x | x | Current HCUP | 13 |
| | Hysterectomy | • HCUP<br>• Submitted as ORYX measure<br>• Literature (Bernstein, Broder) | x | x | Current HCUP | 14 |
| | Joint replacement (hip and knee) | • Literature (Quintana, Van Walraven) | | | Though high variability of procedure makes indicator potentially precise, inappropriateness rates not studied extensively due to inherent subjectivity of indication for procedure. | - |
| | Laminectomy and/or spinal fusion | • HCUP<br>• Submitted as ORYX measure<br>• Literature (Larequi-Lauber, Porchet) | x | x | Current HCUP | 15 |

*An “x” placed in the empirical or literature evidence review column denotes that the review was conducted; a “p” denotes that preliminary analyses were conducted, but that the indicators’ performance was poor on statistical tests of precision.

| Measure Type and Clinical Domain | Indicator Name | Current Users or Developers | Evidence Review* | | Reason for selection or exclusion | Indi-cator # in text body |
|---|---|---|---|---|---|---|
| | | | Empiri-cal | Liter-ature | | |
| | Transurethral resection of the prostate (TURP) | • HCUP<br>• Submitted as ORYX measure | p | | Current HCUP, low precision | - |
| | Radical prostatectomy (all and over 75 years) | • HCUP<br>• Submitted as ORYX measure | p | | Current HCUP, low precision | - |
| | PTCA | • Literature (Bernstein, Leape, Ziskind) | x | x | Common procedure, high area variation, and evidence of misuse. | 16 |
| | Sinus surgery | • Literature (Piccirillo) | | | Potentially outpatient procedure | - |
| | Upper GI tract endoscopy | • Literature (Chassin, Leape) | | | Outpatient procedure | - |
| Length of stay (LOS): | | | | | Since LOS is usually used as a resource measure and not typically as a quality indicator, we did not include any of the LOS indicators in our recommendations. | |
| Medical Examples: | | | | | | |
| | Chemotherapy | • University Hospital Consortium<br>• Submitted as ORYX measure | | | Not typically a quality indicator | - |
| | GI Hemorrhage | • Used in literature as a resource measure | | | Not typically a quality indicator | - |
| | Heart failure and shock | • University Hospital Consortium<br>• Submitted as ORYX measure | | | Not typically a quality indicator | - |
| | HIV related admissions | • University Hospital Consortium | | | Not typically a quality indicator | - |
| | Pneumonia | • University Hospital Consortium<br>• Submitted as ORYX measure | | | Not typically a quality indicator | - |

*An “x” placed in the empirical or literature evidence review column denotes that the review was conducted; a “p” denotes that preliminary analyses were conducted, but that the indicators’ performance was poor on statistical tests of precision.

| Measure Type and Clinical Domain | Indicator Name | Current Users or Developers | Evidence Review* | | Reason for selection or exclusion | Indi-cator # in text body |
|---|---|---|---|---|---|---|
| | | | Empiri-cal | Liter-ature | | |
| | Rehabilitation | • University Hospital Consortium<br>• Submitted as ORYX measure | | | Not typically a quality indicator | - |
| | Stroke | • University Hospital Consortium<br>• Submitted as ORYX measure | | | Not typically a quality indicator | - |
| Surgery Examples: | | | | | | |
| | Bone and joint procedures | • University Hospital Consortium<br>• Submitted as ORYX measure | | | Not typically a quality indicator | - |
| | Coronary procedures | | | | Not typically a quality indicator | - |
| | Circulation disorders with catheterization | | | | Not typically a quality indicator | - |
| | Craniotomy | | | | Not typically a quality indicator | - |
| | Hip replacement surgery | | | | Not typically a quality indicator | - |
| | Percutaneous transluminal coronary angioplasty (PTCA) | | | | Not typically a quality indicator | - |
| Obstetric: | | | | | | |
| | Cesarean section with complications | • University Hospital Consortium | | | Not typically a quality indicator | - |
| | Cesarean section without complications | • University Hospital Consortium<br>• Submitted as ORYX measure | | | Not typically a quality indicator | - |
| | Vaginal delivery with complications | • University Hospital Consortium | | | Not typically a quality indicator | - |
| | Vaginal delivery without complications | • University Hospital Consortium<br>• Submitted as ORYX measure | | | Not typically a quality indicator | - |
| Neonatal: | | | | | | |

*An “x” placed in the empirical or literature evidence review column denotes that the review was conducted; a “p” denotes that preliminary analyses were conducted, but that the indicators’ performance was poor on statistical tests of precision.

| Measure Type and Clinical Domain | Indicator Name | Current Users or Developers | Evidence Review* | | Reason for selection or exclusion | Indi-cator # in text body |
|---|---|---|---|---|---|---|
| | | | Empiri-cal | Liter-ature | | |
| | Normal neonatal | • University Hospital Consortium<br>• Submitted as ORYX measure | | | Not typically a quality indicator | - |
| | Neonatal with complication | • University Hospital Consortium | | | Not typically a quality indicator | - |
| Pediatric: | | | | | | |
| | Pediatric asthma/bronchitis | • University Hospital Consortium<br>• Submitted as ORYX measure | | | Not typically a quality indicator | - |
| Psychiatric: | | | | | | |
| | Psychoses | • University Hospital Consortium<br>• Submitted as ORYX measure | | | Not typically a quality indicator | - |
| Proxy-Outcome measures: | | | | | | |
| Conditional length of stay: | | | | | | |
| | All discharges, length of stay | • HQI ValiData | | | A new indicator in development, will be considered in complications module. | - |
| Outcomes measures: | | | | | | |
| Ambulatory care sensitive conditions/Avoidable hospitalizations: | | | | | | |
| Medical: | | | | | | |
| | AMI | • Literature (Begley) | | | Not included in most commonly used ACSC indicator sets. | - |
| | Cellulitis (or abscess, or other soft tissue infection, or lymphadenitis) | • Billings (United Hospital Fund)<br>• Weissman<br>• Literature (Silver) | p | | Low precision using HCUP NIS data set | - |
| | Congenital syphilis | • Billings (UHF) | | | Infrequent | - |
| | Dehydration/volume depletion | • Billings (UHF) | x | x | Measured with adequate precision. | 17 |

*An "x" placed in the empirical or literature evidence review column denotes that the review was conducted; a "p" denotes that preliminary analyses were conducted, but that the indicators' performance was poor on statistical tests of precision.

| Measure Type and Clinical Domain | Indicator Name | Current Users or Developers | Evidence Review* | | Reason for selection or exclusion | Indi-cator # in text body |
|---|---|---|---|---|---|---|
| | | | Empiri-cal | Liter-ature | | |
| | Delivery, high risk/complicated | • Literature (Laditka, Parchman) | | | Possibly need detailed clinical data. | - |
| | Dental conditions | • Billings (UHF) | | | Infrequent | - |
| | Gangrene | • Weissman | | | Infrequent | - |
| | Gastroenteritis | • Billings (UHF)<br>• Literature (Silver) | p | | Low precision for adults. Evaluated as a pediatric measure. | - |
| | Hypoglycemia | • Billings | p | | Low precision. | |
| | Hypokalemia | • Weissman | | | Infrequent, outside of eating disorder treatment centers | - |
| | Immunization preventable diseases (includes pneumonia and influenza in elderly) | • Billings (UHF)<br>• HCUP<br>• Weissman | p | | Current HCUP Indicator. Low precision, combined with Pneumonia avoidable hospitalization indicator | - |
| | Malnutrition | • Billings (UHF) | | | Infrequent, outside of eating disorder treatment centers | - |
| | Pelvic inflammatory disease | • Billings (UHF) | | | Infrequent | - |
| | Pneumonia (defs may include bacterial pneumonia, pleurisy, bronchitis, bronchiolotis, pharyngitis, and sinsusitis) | • Weissman | x | x | Common admission, measured with adequate precision. | 18 |
| | Pyelonephritis/ Urinary Tract Infection | • UK National Health Service High Level Performance Indicators<br>• Weissman | x | x | Measured with adequate precision | 19 |
| | Ruptured/perforated appendix | • HCUP<br>• Weissman | x | x | Current HCUP QI, measured with adequate precision. | 20 |
| | Severe ear, nose and throat infections | • Billings (UHF) | p | | Low precision using HCUP NIS data set | - |
| | Skin graft with cellulitis | • Billings | p | | Low precision using HCUP NIS data set | |

*An “x” placed in the empirical or literature evidence review column denotes that the review was conducted; a “p” denotes that preliminary analyses were conducted, but that the indicators’ performance was poor on statistical tests of precision.

| Measure Type and Clinical Domain | Indicator Name | Current Users or Developers | Evidence Review* | | Reason for selection or exclusion | Indi-cator # in text body |
|---|---|---|---|---|---|---|
| | | | Empiri-cal | Liter-ature | | |
| | Stroke (def may specify with hypertension) | • Literature (Begley, Djojonegoro) | | | Not included in most commonly used ACSC indicator sets. | - |
| | Tuberculosis (may include other respiratory infections) | • Billings (UHF)<br>• Literature (Rohrer) | | | Infrequent | - |
| Chronic: | | | | | | |
| | Angina | • Billings<br>• Literature (Rohrer) | x | x | Common admission, measured with adequate precision. | 21 |
| | Anemia (includes iron deficiency anemia) | • Billings<br>• Literature (Silver) | | | | |
| | Asthma | • Billings (UHF)<br>• UK National Health Service High Level Performance Indicators<br>• HP 2010<br>• Weissman | x | x | A current HCUP QI as a pediatric measure. | 22 |
| | Asthma and bronchitis | • Literature (Rohrer) | | | Evaluated asthma without bronchitis. | - |
| | Cancer, breast (female) | • Literature (Shukla) | | | Not included in most sets of ACSC indicators. | - |
| | Cancer, cervical (invasive) | • Literature (Shukla) | | | Not included in most sets of ACSC indicators. | - |
| | Cerebrovascular disease among non-elderly adults | • HCUP | p | | Current HCUP QI, low precision using HCUP NIS data set. | - |
| | Chronic obstructive pulmonary disease (COPD) | • Billings (UHF)<br>• Colorado Health and Hospital, proposed | x | x | Frequent cause of admission, guidelines for ambulatory management, differences in practice patterns. | 23 |

*An “x” placed in the empirical or literature evidence review column denotes that the review was conducted; a “p” denotes that preliminary analyses were conducted, but that the indicators’ performance was poor on statistical tests of precision.

| Measure Type and Clinical Domain | Indicator Name | Current Users or Developers | Evidence Review* | | Reason for selection or exclusion | Indi-cator # in text body |
|---|---|---|---|---|---|---|
| | | | Empiri-cal | Liter-ature | | |
| | Congestive heart failure (CHF; may include CHF and pulmonary edema or shock) | • Billings (UHF)<br>• DEMPAQ (CHF complications)<br>• UK National Health Service High Level Performance Indicators<br>• Weissman<br>• Literature (Rohrer) | x | x | Common chronic condition with relatively frequent hospitalizations | 24 |
| | Diabetes, all | • Literature (Connell) | | | | |
| | Diabetes (short-term complications or A,B,C) | • Billings (UHF)<br>• DEMPAQ<br>• HCUP<br>• UK National Health Service High Level Performance Indicators (single diabetes measure)<br>• Weissman | x | x | Current HCUP QI with complications | 25 |
| | Diabetes, uncontrolled | • HP 2010<br>• Literature (Bierman, Shukla) | x | x | Addendum indicator for use with diabetes short-term complications to make definition consistent with HP2010. | 26 |
| | Diabetes (long-term complications) | • HCUP<br>• UK National Health Service High Level Performance Indicators (single diabetes measure)<br>• Literature (Begley) | x | x | Current HCUP QI with complications | 27 |
| | Diabetic ketoacidosis (DKA) | • DEMPAQ<br>• Sagamore Health<br>• Weissman | | | A current HCUP QI as a diabetes complication measure | - |

*An “x” placed in the empirical or literature evidence review column denotes that the review was conducted; a “p” denotes that preliminary analyses were conducted, but that the indicators’ performance was poor on statistical tests of precision.

| Measure Type and Clinical Domain | Indicator Name | Current Users or Developers | Evidence Review* | | Reason for selection or exclusion | Indi-cator # in text body |
|---|---|---|---|---|---|---|
| | | | Empiri-cal | Liter-ature | | |
| | Hypertension (includes malignant) | • Billings (UHF)<br>• Weissman | x | x | Measured with adequate precision. | 28 |
| | Lower extremity amputation | • DEMPAQ<br>• HP 2010 | x | x | Healthy People 2000 goal, low precision using HCUP NIS data set. | 29 |
| | Perforated/bleeding ulcer | • Weissman | | | Infrequent | - |
| | Rheumatic fever | • Literature (Shukla) | | | | |
| | Seizures and convulsions (may include convulsions 'A' & 'B'; epilepsy or grand mal status) | • Billings (UHF)<br>• UK National Health Service High Level Performance Indicators | p | | Low precision | - |
| Pediatrics: | | | | | | |
| | Failure to thrive | • Billings (UHF) | | | Infrequent | - |
| | Immunization preventable diseases (measles, mumps, and/or polio) | • Billings<br>• South Carolina adaptation of Billings (UHF)<br>• Weissman<br>• Literature (Shukla) | | | Infrequent | - |
| | Iron deficiency anemia | • Billings (UHF) | | | Infrequent | - |
| | Low birthweight | • HCUP<br>• HEDIS<br>• United Health Care<br>• University Hospital Consortium<br>• Literature (Shukla) | x | x | Current HCUP QI | 30 |
| | Pediatric acute otitis media | • South Carolina adaptation of Billings (UHF)<br>• Literature (Rohrer) | | | Infrequent (as a primary diagnosis excluding surgery patients) | - |

*An “x” placed in the empirical or literature evidence review column denotes that the review was conducted; a “p” denotes that preliminary analyses were conducted, but that the indicators’ performance was poor on statistical tests of precision.

| Measure Type and Clinical Domain | Indicator Name | Current Users or Developers | Evidence Review* | | Reason for selection or exclusion | Indi-cator # in text body |
|---|---|---|---|---|---|---|
| | | | Empiri-cal | Liter-ature | | |
| | Pediatric Asthma (may include bronchitis & asthma) | • Billings<br>• Colorado Health and Hospital, proposed<br>• HCUP<br>• South Carolina adaptation of Billings (UHF)<br>• United Health Care<br>• Weissman<br>• Literature (Rohrer) | x | x | Frequent cause of admission, guidelines for ambulatory management, differences in practice patterns | 31 |
| | Pediatric burns | • Literature (Gadomski et al.) | | | A good potential injury indicator, which was not our primary focus. | - |
| | Pediatric diabetes | • Literature (Gadomski et al.) | p | | Frequent cause of admission, very low precision | - |
| | Pediatric gastroenteritis | • South Carolina adaptation of Billings (UHF) | x | x | Frequent cause of admission, guidelines for ambulatory management, differences in practice patterns | 32 |
| | Pediatric mastoiditis | • Literature | | | Infrequent | - |
| | Pediatric nausea and vomiting | • Literature (Gadomski et al.) | | | Infrequent | - |
| | Pediatric urinary tract infection | • South Carolina adaptation of Billings (UHF) | | | Infrequent | - |
| | Pediatric viral meningitis | • Literature (Gadomski et al.) | | | Infrequent | |
| | Pediatric viral syndrome | • Literature (Gadomski et al.) | | | Frequent cause of admission | - |
| | Very low birthweight | • HEDIS<br>• IMSystem<br>• HCUP | p | | Current HCUP, low precision using HCUP NIS data set. | - |
| Complications: | | | | | REVIEW OF COMPLICATIONS IN SEPARATE REPORT | |
| In-hospital Mortality: | | | | | | |

*An “x” placed in the empirical or literature evidence review column denotes that the review was conducted; a “p” denotes that preliminary analyses were conducted, but that the indicators’ performance was poor on statistical tests of precision.

| Measure Type and Clinical Domain | Indicator Name | Current Users or Developers | Evidence Review* | | Reason for selection or exclusion | Indi-cator # in text body |
|---|---|---|---|---|---|---|
| | | | Empiri-cal | Liter-ature | | |
| Medical: | Overall mortality | • Collaborative Approach to Resource Effectiveness (CARE) Project<br>• Maryland QI Project<br>• Literature<br>• Submitted as ORYX measure | | | Potentially very biased due to heterogeneous patient population; difficult to determine an appropriate intervention | - |

*An “x” placed in the empirical or literature evidence review column denotes that the review was conducted; a “p” denotes that preliminary analyses were conducted, but that the indicators’ performance was poor on statistical tests of precision.

| Measure Type and Clinical Domain | Indicator Name | Current Users or Developers | Evidence Review* | | Reason for selection or exclusion | Indi-cator # in text body |
|---|---|---|---|---|---|---|
| | | | Empiri-cal | Liter-ature | | |
| | Acute myocardial infarction (AMI) | • California Hospital Outcomes Project<br>• Cleveland Health Quality Choice (grouped with cardiovascular measure)<br>• Collaborative Approach to Resource Effectiveness (CARE) Project<br>• Colorado Health and Hospital, proposed<br>• Greater New York Hospital Association<br>• HealthGrades.com<br>• IMSystem<br>• JCAHO core measure<br>• Michigan Hospital Association (in aggregated measure)<br>• Pennsylvania Health Care Cost Containment Council<br>• University Hospital Consortium<br>• UK National Health Service High Level Performance Indicators<br>• Submitted as ORYX measure | x | x | Already a substantial amount of information on hospital performance with AMI, including the Medicare Cooperative Cardiovascular project; other HCUP QI focus on procedures for the treatment of MI (PTCA, CABG). | 33 |

*An “x” placed in the empirical or literature evidence review column denotes that the review was conducted; a “p” denotes that preliminary analyses were conducted, but that the indicators’ performance was poor on statistical tests of precision.

| Measure Type and Clinical Domain | Indicator Name | Current Users or Developers | Evidence Review* | | Reason for selection or exclusion | Indi-cator # in text body |
|---|---|---|---|---|---|---|
| | | | Empiri-cal | Liter-ature | | |
| | Angina | • Michigan Hospital Association (in aggregated measure) | | | Have indicator for AMI mortality, one of the other conditions covered by the MHA aggregate measure | - |
| | Bone marrow transplant (BMT) | • University Hospital Consortium | | | Limited number of providers, infrequent | - |
| | Chronic obstructive pulmonary disease | • Literature (Rosenthal et al.)<br>• Cleveland Health Quality Choice (grouped with respiratory measure)<br>• HealthGrades.com<br>• Maryland QI Project<br>• Michigan Hospital Association (in aggregated measure)<br>• Submitted as ORYX measure | | | A progressive chronic disease; more difficult to evaluate without longitudinal data | - |

*An "x" placed in the empirical or literature evidence review column denotes that the review was conducted; a "p" denotes that preliminary analyses were conducted, but that the indicators' performance was poor on statistical tests of precision.

| Measure Type and Clinical Domain | Indicator Name | Current Users or Developers | Evidence Review* | | Reason for selection or exclusion | Indi-cator # in text body |
|---|---|---|---|---|---|---|
| | | | Empiri-cal | Liter-ature | | |
| | Congestive heart failure (CHF) | • Literature (Rosenthal et al.)<br>• Cleveland Health Quality Choice (grouped with cardiovascular measure)<br>• HealthGrades.com<br>• Greater New York Hospital Association<br>• Maryland QI Project<br>• Michigan Hospital Association (in aggregated measure)<br>• University Hospital Consortium<br>• Submitted as ORYX measure | x | x | Substantial literature, some use as quality indicator. Related to other indicators as overall cardiovascular care. | 34 |
| | Adult diabetes | • Pennsylvania Health Care Cost Containment Council | | | Very rare cause of death ; other indicators (cardiovascular, infections) capture deaths in this population. | - |
| | GI hemorrhage | • Literature (Rosenthal et al.)<br>• Cleveland Health Quality Choice<br>• Maryland QI Project<br>• Michigan Hospital Association (in aggregated measure)<br>• Outcome measure in literature | x | x | Indicator selected to focus on related aspects of particular condition (i.e., mortality, utilization, volume). | 35 |

*An “x” placed in the empirical or literature evidence review column denotes that the review was conducted; a “p” denotes that preliminary analyses were conducted, but that the indicators’ performance was poor on statistical tests of precision.

| Measure Type and Clinical Domain | Indicator Name | Current Users or Developers | Evidence Review* | | Reason for selection or exclusion | Indi-cator # in text body |
|---|---|---|---|---|---|---|
| | | | Empiri-cal | Liter-ature | | |
| | Heart failure and shock | • Pennsylvania Health Care Cost Containment Council<br>• University Hospital Consortium<br>• Submitted as ORYX measure | p | | Indicator selected to focus on related aspects of particular condition (i.e., mortality, utilization, volume), measured with low precision. | - |
| | Hip fracture | • Literature<br>• UK National Health Service High Level Performance Indicators | x | x | Common admission in the elderly with relatively high mortality rates. Substantial literature regarding hip fracture mortality. | 36 |
| | HIV | • Maryland QI Project | | | Data unavailable. Confounds due to HAART, home IV antibiotics, etc. | - |
| | Lung cancer | • Pennsylvania Health Care Cost Containment Council | | | A progressive disease requiring clinical risk adjustment (PHC4 uses clinical risk adjustment) | |

*An “x” placed in the empirical or literature evidence review column denotes that the review was conducted; a “p” denotes that preliminary analyses were conducted, but that the indicators’ performance was poor on statistical tests of precision.

| Measure Type and Clinical Domain | Indicator Name | Current Users or Developers | Evidence Review* | | Reason for selection or exclusion | Indi-cator # in text body |
|---|---|---|---|---|---|---|
| | | | Empiri-cal | Liter-ature | | |
| | Pneumonia | • Literature (Rosenthal et al.)<br>• California Hospital Outcomes Project<br>• Greater New York Hospital Association<br>• HealthGrades.com<br>• Maryland QI Project<br>• Michigan Hospital Association (in aggregated measure)<br>• Pennsylvania Health Care Cost Containment Council<br>• University Hospital Consortium<br>• Submitted as ORYX measure | x | x | Indicator selected to focus on related aspects of particular condition (i.e., mortality, utilization, volume). | 37 |
| | Pneumonia/hemo thorax mortality | • IMSystem | | | Limited use | - |
| | Renal failure | • Maryland Hospital Association<br>• Pennsylvania Health Care Cost Containment Council | | | Disease is rarely fatal by itself, associated with other serious complications or comorbidities. | - |
| | Respiratory failure | • Maryland Hospital Association | | | Often coded based on physiological state, rather than as a treatable condition. | - |
| | Septicemia | • Maryland Hospital Association<br>• Pennsylvania Health Care Cost Containment Council | | | Limited evidence located. | - |

*An “x” placed in the empirical or literature evidence review column denotes that the review was conducted; a “p” denotes that preliminary analyses were conducted, but that the indicators’ performance was poor on statistical tests of precision.

| Measure Type and Clinical Domain | Indicator Name | Current Users or Developers | Evidence Review* | | Reason for selection or exclusion | Indi-cator # in text body |
|---|---|---|---|---|---|---|
| | | | Empiri-cal | Liter-ature | | |
| | Stroke | • Literature (Rosenthal et al.)<br>• Cleveland Health Quality Choice<br>• Greater New York Hospital Association<br>• HealthGrades.com<br>• Maryland QI Project<br>• Michigan Hospital Association (in aggregated measure)<br>• Submitted as ORYX measure | x | x | Indicator selected to focus on related aspects of particular condition (i.e., mortality, utilization, volume). | 38 |
| | Trauma | • Literature, Rutledge, Hannan | | | Potential need for emergency room data and/or clinical data on injury severity. | - |
| Surgical: | | | | | | |
| | Abdominal aortic aneurysm repair | • Literature (volume-outcome)<br>• HealthGrades.com<br>• Pennsylvania Health Care Cost Containment Council (with major vessel procedures) | x | x | Indicator selected to focus on related aspects of particular condition (i.e., mortality, utilization, volume). | 39 |
| | Carotid endarterectomy | • Literature (volume-outcome)<br>• Greater New York Hospital Association<br>• Pennsylvania Health Care Cost Containment Council (with vascular operations) | p | | Preliminary exploration of indicator to focus on related aspects of particular condition (i.e., mortality, utilization, volume). Excluded due to low precision. | - |

*An "x" placed in the empirical or literature evidence review column denotes that the review was conducted; a "p" denotes that preliminary analyses were conducted, but that the indicators' performance was poor on statistical tests of precision.

| Measure Type and Clinical Domain | Indicator Name | Current Users or Developers | Evidence Review* | | Reason for selection or exclusion | Indi-cator # in text body |
|---|---|---|---|---|---|---|
| | | | Empiri-cal | Liter-ature | | |
| | Circulation disorder with catheterization | • University Hospital Consortium | | | In general, avoid non-complication related procedure based mortality measures due to problems of patient selection | - |
| | Cholecystectomy | • HCUP<br>• Submitted as ORYX measure | p | | Current HCUP QI, low precision. | - |
| | Coronary artery bypass graft (CABG) | • California CABG Mortality Reporting Project<br>• Greater New York Hospital Association<br>• HealthGrades.com<br>• IMSystem<br>• Maryland QI Project<br>• New Jersey Depart. of Health and Senior Services<br>• New York Dept. of Health<br>• Pennsylvania Health Care Cost Containment Council<br>• University Hospital Consortium<br>• Submitted as ORYX measure | x | x | Already widely reported in multiple states. | 40 |
| | Coronary Procedures | • University Hospital Consortium | | | In general, avoid non-complication related procedure based mortality measures due to problems of patient selection | - |

*An "x" placed in the empirical or literature evidence review column denotes that the review was conducted; a "p" denotes that preliminary analyses were conducted, but that the indicators' performance was poor on statistical tests of precision.

| Measure Type and Clinical Domain | Indicator Name | Current Users or Developers | Evidence Review* | | Reason for selection or exclusion | Indi-cator # in text body |
|---|---|---|---|---|---|---|
| | | | Empiri-cal | Liter-ature | | |
| | Craniotomy | • University Hospital Consortium<br>• Submitted as ORYX measure | x | x | Indicator selected to focus on related aspects of particular condition (i.e., mortality, utilization, volume) | 41 |
| | Esophageal resection | • Literature (volume - outcome) | x | x | Indicator selected to focus on related aspects of particular condition (i.e., mortality, utilization, volume) | 42 |
| | Knee replacement | • HCUP<br>• Greater New York Hospital Association<br>• HealthGrades.com<br>• Medicare Quality of care Surveillance System<br>• University Hospital Consortium (aggregated with THA)<br>• Submitted as ORYX measure | p | | Current HCUP QI. Low precision. | - |

*An “x” placed in the empirical or literature evidence review column denotes that the review was conducted; a “p” denotes that preliminary analyses were conducted, but that the indicators’ performance was poor on statistical tests of precision.

| Measure Type and Clinical Domain | Indicator Name | Current Users or Developers | Evidence Review* | | Reason for selection or exclusion | Indi-cator # in text body |
|---|---|---|---|---|---|---|
| | | | Empiri-cal | Liter-ature | | |
| | Hip replacement | • HCUP<br>• Greater New York Hospital Association<br>• HealthGrades.com<br>• Medicare Quality of care Surveillance System<br>• Pennsylvania Health Care Cost Containment Council (with hip operations)<br>• University Hospital Consortium (aggregated with TKA)<br>• Submitted as ORYX measure | x | x | Current HCUP QI | 43 |
| | Hysterectomy | • HCUP<br>• Medicare Quality of care Surveillance System<br>• Submitted as ORYX measure | p | | Current HCUP QI. Low precision. | - |
| | Laminectomy/spinal fusion | • HCUP<br>• HealthGrades.com<br>• Medicare Quality of care Surveillance System<br>• Submitted as ORYX measure | p | | Current HCUP QI. Low precision. | - |
| | Orthopedic surgery | • University Hospital Consortium<br>• Submitted as ORYX measure | | | In general, avoid non-complication related procedure based mortality measures due to problems of patient selection | - |

*An “x” placed in the empirical or literature evidence review column denotes that the review was conducted; a “p” denotes that preliminary analyses were conducted, but that the indicators’ performance was poor on statistical tests of precision.

| Measure Type and Clinical Domain | Indicator Name | Current Users or Developers | Evidence Review* | | Reason for selection or exclusion | Indi-cator # in text body |
|---|---|---|---|---|---|---|
| | | | Empiri-cal | Liter-ature | | |
| | Pancreatic resection | • Literature (volume-outcome) | x | x | Indicator selected to focus on related aspects of particular condition (i.e., mortality, utilization, volume) | 44 |
| | Perioperative mortality | • Collaborative Approach to Resource Effectiveness (CARE) Project<br>• IMSystem<br>• Maryland QI Project<br>• University Hospital Consortium<br>• Michigan Hospital Association (with specified procedures)<br>• Submitted as ORYX measure | | | Potentially very biased due to heterogeneous patient population; difficult to determine an appropriate intervention | - |
| | Percutaneous transluminal coronary angioplasty (PTCA) | • IM System<br>• Greater New York Hospital Association<br>• HealthGrades.com<br>• New York Dept. of Health<br>• University Hospital Consortium<br>• Submitted as ORYX measure | p | | Indicator selected to focus on related aspects of particular condition (i.e., mortality, utilization, volume) Low precision. | - |

*An “x” placed in the empirical or literature evidence review column denotes that the review was conducted; a “p” denotes that preliminary analyses were conducted, but that the indicators’ performance was poor on statistical tests of precision.

| Measure Type and Clinical Domain | Indicator Name | Current Users or Developers | Evidence Review* | | Reason for selection or exclusion | Indi-cator # in text body |
|---|---|---|---|---|---|---|
| | | | Empiri-cal | Liter-ature | | |
| | Prostatectomy | • Literature (Klein et al.)<br>• Medicare Quality of care Surveillance System<br>• Submitted as ORYX measure | | | In general, avoid non-complication related procedure based mortality measures due to problems of patient selection | - |
| | Transurethral resection of prostate (TURP) | • HCUP<br>• Submitted as ORYX measure | p | x | Current HCUP QI. Low precision. | - |
| Pediatric: | | | | | | |
| | Overall pediatric mortality | • University Hospital Consortium | | | Potentially very biased due to heterogeneous patient population; difficult to determine an appropriate intervention | - |
| | Neonatal mortality | • Collaborative Approach to Resource Effectiveness (CARE) Project<br>• Maryland QI Project<br>• University Hospital Consortium<br>• Submitted as ORYX measure | | | Potentially very biased due to heterogeneous patient population; difficult to determine an appropriate intervention | - |
| | Pediatric heart surgery | • Literature (volume-outcome) | x | x | Indicator selected to focus on related aspects of particular condition (i.e., mortality, utilization, volume) | 45 |
| Psychiatric: | | | | | | |
| | Depression | • RAND, JCAHO | | | Infrequent in acute care setting | - |
| Other measures: | | | | | | |
| | Newborn Outcomes | • IMSystem | | | Difficult to implement without longitudinal data | - |

*An "x" placed in the empirical or literature evidence review column denotes that the review was conducted; a "p" denotes that preliminary analyses were conducted, but that the indicators' performance was poor on statistical tests of precision.

| Measure Type and Clinical Domain | Indicator Name | Current Users or Developers | Evidence Review* | | Reason for selection or exclusion | Indi-cator # in text body |
|---|---|---|---|---|---|---|
| | | | Empiri-cal | Liter-ature | | |
| Public Health Measures: | | | | | | |
| | Admission for AMI | • Colorado Health and Hospital, proposed | | | Included in previously considered indicators | - |
| | Adolescent mothers | • Colorado Health and Hospital, proposed | | | These indicators generally are underreported, and have less connection with the health care system | - |
| | Child abuse | • Colorado Health and Hospital, proposed | | | These indicators generally are underreported, and have less connection with the health care system | - |
| | External injury (firearms) | • Colorado Health and Hospital, proposed | | | These indicators generally are underreported, and have less connection with the health care system | - |
| | Motor vehicle accident (MVA) | • Colorado Health and Hospital, proposed | | | These indicators generally are underreported, and have less connection with the health care system | - |
| | Substance abuse | • Colorado Health and Hospital, proposed | | | Difficult to determine from discharge data true cases, infrequent in acute care setting | - |
| | Stroke/TIA | • Colorado Health and Hospital, proposed | | | Included in previously considered indicators | - |
| | Suicide attempt | • Colorado Health and Hospital, proposed | | | Concern about reporting bias. | - |

*An “x” placed in the empirical or literature evidence review column denotes that the review was conducted; a “p” denotes that preliminary analyses were conducted, but that the indicators’ performance was poor on statistical tests of precision.

# APPENDIX 6

# Indicator Definitions

Table of indicator includes complete specification of current HCUP indicators and potential HCUP indicators evaluated in this evidence report.

Indicators are organized by indicator type. Each indicator is numbered with a unique identifier used throughout the report.

The numerator refers to the outcome of interest and the denominator refers to the population at risk.

Some indicators specification are so detailed that the full specifications are included after the table.

## Appendix 6 Indicator Definitions – Potential HCUP Indicators.

** indicates current HCUP indicators

| Indicator | Numerator | Denominator | # |
|---|---|---|---|
| **Volume Outcome** | | | |
| ▪ AAA repair volume | Discharges with ICD-9 codes 38.34, 38.44, or 38.64 in any procedure fie ld and diagnosis of AAA (see below) in any field.<br><br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | Not applicable. | 1 |
| ▪ Carotid Endarterectomy volume | Discharges with ICD-9 code 38.12 in any procedure field.<br><br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | Not applicable | 2 |
| ▪ CABG volume | Discharges with ICD-9 codes 36.10 – 36.19 in any procedure field.<br><br>Age 40 years and older.<br><br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | Not applicable | 3 |
| ▪ Esophageal resection volume | Discharges with ICD-9 codes 42.40 – 42.42 in any procedure field and diagnosis code of esophageal cancer in any field.<br><br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | Not applicable | 4 |
| ▪ Pancreatic resection volume | Discharges with ICD-9 codes 52.6 or 52.7 in any procedure field and diagnosis code of pancreatic cancer in any field.<br><br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other | Not applicable | 5 |

| Indicator | Numerator | Denominator | # |
|---|---|---|---|
| | neonates). | | |
| ▪ Pediatric heart surgery volume | Discharges with ICD-9 codes for specified heart surgery (see below) in any field, or for any heart surgery and a diagnosis of hypoplastic left heart syndrome (see below) in any field<br><br>Age less than 18.<br>See NOTE for additional exclusions<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium). | Not applicable | 6 |
| ▪ PTCA volume | Discharges with ICD-9 codes 36.01, 36.02, 36.05, or 36.06 in any procedure field.<br><br>Age 40 years and older.<br><br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | Not applicable | 7 |
| **Utilization – provider level** | | | |
| ▪ Cesarean section delivery rate | Number of C-sections per 100 deliveries (see below). | All deliveries (see below) | 8 |
| ▪ Incidental appendectomy among elderly rate | Number of incidental appendectomies per 100 elderly discharges with intra-abdominal procedure (see below). | All non-maternal/non-neonatal discharges age 65 years and older with intra-abdominal procedure in any procedure field (see below).<br><br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | 9 |
| ▪ Bi-lateral catheterization rate | Number of simultaneous right and left heart catheterizations per 100 discharges with procedure code of heart catheterization (see below).<br><br>Exclude valid indications for right sided catheterization (see below) in any diagnosis field. | All non-maternal/non-neonatal discharges with heart catheterization (see below) in any procedure field.<br><br>Include only coronary artery disease (see below)<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | 10 |
| ▪ Successful vaginal birth after cesarean section (VBAC) rate | Number of vaginal births per 100 deliveries with diagnosis of previous C-section (see below). | All deliveries with previous C-section diagnosis in any diagnosis field (see below). | 11 |

| Indicator | Numerator | Denominator | # |
|---|---|---|---|
| ▪ Laparoscopic cholecystectomy rate | Number of laparoscopic cholecystectomies per 100 discharges with procedure code of cholecystectomy (see below). | All non-maternal/non-neonatal discharges with cholecystectomy (see below) in any procedure field.<br><br>Include only discharges with uncomplicated cases: cholecystitis and/or cholelithiasis (see below) in any diagnosis field.<br><br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | 12 |
| **Utilization – area level** | | | |
| ▪ Coronary artery bypass graft (CABG) rate | Number of CABGs (any procedure field) per 100,000 population (see below).<br><br>All non-maternal/non-neonatal discharges age 40 years and older.<br><br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | Population in MSA or county, age 40 years and older. | 13 |
| ▪ Hysterectomy rate | Number of hysterectomies (any procedure field) per 100,000 population (see below).<br><br>All non-maternal/non-neonatal discharges of females age 18 years and older.<br><br>Exclude discharges with genital cancer or pelvic or lower abdominal trauma (see below) in any diagnosis field.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | Female population in MSA or county, age 18 years and older. | 14 |
| ▪ Laminectomy and/or spinal fusion rate | Number of laminectomies and/or spinal fusions (any procedure field) per 100,000 population (see below).<br><br>All non-maternal/non-neonatal discharges of age 18 years and older. | Population in MSA or county, age 18 years and older. | 15 |

| Indicator | Numerator | Denominator | # |
|---|---|---|---|
| | Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | | |
| ▪ PTCA rate | Number of PTCAs (any procedure field) per 100,000 population (see below).<br><br>All non-maternal/non-neonatal discharges of age 40 years and older.<br><br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | Population in MSA or county, age 40 years and older. | 16 |
| **Ambulatory Care Sensitive Conditions** | | | |
| ▪ Dehydration admission rate | Discharges with ICD-9 principal diagnosis code for hypovolemia per 100,000 population (see below).<br><br>All non-maternal/non-neonatal discharges of age less than 65*.<br><br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | Population in MSA or county, age less than 65*.<br><br>* Rate can also be calculated for age 65 and older | 17 |
| ▪ Bacterial pneumonia admission rate | Discharges with ICD-9 principal diagnosis code for bacterial pneumonia per 100,000 population.<br><br>All non-maternal/non-neonatal discharges of age less than 65*.<br><br>Exclude discharges with diagnosis code for sickle cell anemia or HB-S disease (see below) in any field.<br>Exclude patients <2 months (8 weeks) of age<br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | Population in MSA or county, age less than 65*.<br><br>* Rate can also be calculated for age 65 and older. | 18 |
| ▪ Urinary infection admission rate | Discharges with ICD-9 principal diagnosis code of | Population in MSA or county, age less than 65*. | 19 |

| Indicator | Numerator | Denominator | # |
|---|---|---|---|
| | urinary tract infection per 100,000 population (see below).<br><br>All non-maternal/non-neonatal discharges of age less than 65*.<br><br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | * Rate can also be calculated for age 65 and older. | |
| ▪ Perforated appendix admission rate | Discharges with ICD-9 diagnosis code for perforations or abscesses of appendix (see below) in any field per 100 discharges with diagnosis code for appendicitis (see below).<br><br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | Number of discharges with diagnosis code for appendicitis in any field in MSA or county. | 20 |
| ▪ Angina admission rate | Discharges with ICD-9 principal diagnosis code for angina (see below) per 100,000 population.<br><br>All non-maternal/non-neonatal discharges of age 18 to 64 years old*.<br><br>Exclude discharges with a surgical procedure in any field (01.0-86.99).<br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | Population in MSA or county, age 18 to 64 years old*.<br><br>* Rate can also be calculated for age 65 and older. | 21 |
| ▪ Adult asthma admission rate | Discharges with ICD-9 principal diagnosis code of asthma per 100,000 population (see below).<br><br>All non-maternal/non-neonatal discharges of age 18 to 64 years old*.<br><br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and | Population in MSA or county, age 18 to 64 years old*.<br><br>* Rate can also be calculated for age 65 and older. | 22 |

| Indicator | Numerator | Denominator | # |
|---|---|---|---|
| | puerperium) and MDC 15 (newborns and other neonates). | | |
| ▪ Chronic Obstructive Pulmonary Disease admission rate | Discharges with ICD-9 principal diagnosis code for COPD (see below) per 100,000 population.<br><br>All non-maternal/non-neonatal discharges of age 18 to 64 years old*.<br><br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | Population in MSA or county, age 18 to 64 years old*.<br><br>* Rate can also be calculated for age 65 and older. | 23 |
| ▪ Congestive Heart Failure admission rate | Discharges with ICD-9 principal diagnosis code for CHF (see below) per 100,000 population.<br><br>All non-maternal/non-neonatal discharges of age 18 to 64 years old*.<br><br>Exclude discharges with cardiac procedure codes (see below) in any field.<br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | Population in MSA or county, age 18 to 64 years old*.<br><br>* Rate can also be calculated for age 65 and older. | 24 |
| ▪ Diabetes – short-term complications admission rate | Discharges with ICD-9 principal diagnosis code for short-term complications (ketoacidosis, hyperosmolarity, coma) per 100,000 population (see below).<br><br>All non-maternal/non-neonatal discharges of age 18 to 64 years old*.<br><br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | Population in MSA or county, age 18 to 64 years old*.<br><br>* Rate can also be calculated for age 65 and older. | 25 |
| ▪ Diabetes – uncontrolled without complications admission rate | Discharges with ICD-9 principal diagnosis code for uncontrolled diabetes, without mention of a short-term or long-term complication, per 100,000 | Population in MSA or county, age 18 to 64 years old*. | 26 |

| Indicator | Numerator | Denominator | # |
|---|---|---|---|
| MAY BE COMBINED WITH DIABETES SHORT-TERM COMPLICATIONS AS A SINGLE INDICATOR. | population (see below).<br><br>All non-maternal/non-neonatal discharges of age 18 to 64 years old*.<br><br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | * Rate can also be calculated for age 65 and older. | |
| ▪ Diabetes – long-term complications admission rate | Discharges with ICD-9 principal diagnosis code for long-term complications (renal, eye, neurological, circulatory, or complications not otherwise specified) per 100,000 population (see below).<br><br>All non-maternal/non-neonatal discharges of age 18 to 64 years old*.<br><br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | Population in MSA or county, age 18 to 64 years old*.<br><br>* Rate can also be calculated for age 65 and older. | 27 |
| ▪ Hypertension admission rate | Discharges with ICD-9 principal diagnosis code for hypertension per 100,000 population (see below).<br><br>All non-maternal/non-neonatal discharges of age 18 to 64 years old*.<br><br>Exclude discharges with cardiac procedure codes (see below) in any field.<br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | Population in MSA or county, age 18 to 64 years old*.<br><br>* Rate can also be calculated for age 65 and older. | 28 |
| ▪ Lower extremity amputation in diabetics admission rate | Discharges with ICD-9 procedure code for lower extremity amputation (see below) in any field and diagnosis code of diabetes in any field per 100,000 population (see below).<br><br>All non-maternal/non-neonatal discharges of age 18 | Population in MSA or county, age 18 to 64 years old*.<br><br>* Rate can also be calculated for age 65 and older. | 29 |

| Indicator | Numerator | Denominator | # |
|---|---|---|---|
| | to 64 years old*.<br><br>Exclude trauma (see below).<br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | | |
| ▪ Low birthweight rate | Number of births with ICD-9 diagnosis code for less than 2500 grams per 100 births (see below) in any field.<br><br>Exclude transfer from other institution. | All births (discharges in MDC 15 - newborns and other neonates) in MSA or county. | 30 |
| ▪ Pediatric asthma admission rate | Discharges with ICD-9 principal diagnosis code of asthma per 100,000 population (see below).<br><br>All non-maternal/non-neonatal discharges of under age 18.<br><br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | Population in MSA or county, under age 18. | 31 |
| ▪ Pediatric gastroenteritis admission rate | Discharges with ICD-9 principal diagnosis code for gastroenteritis per 100,000 population (see below).<br><br>All non-maternal/non-neonatal discharges of under age 18.<br><br>Exclude transfer from other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | Population in MSA or county, under age 18. | 32 |
| **In-hospital Mortality** | | | |
| ▪ Acute myocardial infarction mortality rate | Number of deaths per 100 discharges with diagnosis code for AMI in any field. | All non-maternal/non-neonatal discharges with diagnosis codes for AMI (see below) in any field, age 18 years and older.<br><br>Exclude transfers to other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and | 33 |

| Indicator | Numerator | Denominator | # |
|---|---|---|---|
| | | puerperium) and MDC 15 (newborns and other neonates). | |
| ▪ Congestive Heart Failure mortality rate | Number of deaths per 100 discharges with principal diagnosis code for CHF. | All non-maternal/non-neonatal discharges with principal diagnostic code for CHF (see below) , age 18 years and older.<br><br>Exclude discharges with cardiac procedure codes (see below) in any field<br>Exclude transfers to other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | 34 |
| ▪ GI Hemorrhage mortality rate | Number of deaths per 100 discharges with principal diagnosis code for GI hemorrhage. | All non-maternal/non-neonatal discharges with principal diagnostic code for gastrointestinal hemorrhage (see below), age 18 years and older.<br><br>Exclude transfers to other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates | 35 |
| ▪ Hip fracture mortality rate | Number of deaths per 100 discharges with principal diagnosis code for hip fracture. | All non-maternal/non-neonatal discharges with principal diagnostic code for hip fracture (see below) , age 18 years and older.<br><br>Exclude transfers to other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | 36 |
| ▪ Pneumonia mortality rate | Number of deaths per 100 discharges with principal diagnosis code for pneumonia. | All non-maternal/non-neonatal discharges with principal diagnostic code for pneumonia (see below) , age 18 years and older.<br><br>Exclude transfers to other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates | 37 |
| ▪ Stroke mortality rate | Number of deaths per 100 discharges with principal diagnosis code for stroke. | All non-maternal/non-neonatal discharges with principal diagnostic code for stroke (see below) , age 18 years and older. | 38 |

| Indicator | Numerator | Denominator | # |
|---|---|---|---|
| | | Exclude transfers to other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates | |
| **Post-procedural Mortality** | | | |
| ▪ Abdominal aortic artery repair mortality rate | Number of deaths per 100 discharges with procedure code for AAA repair and diagnosis of AAA in any field. | All non-maternal/non-neonatal discharges with procedure code for AAA repair and diagnosis of AAA (see below) in any field.<br><br>Exclude transfers to other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | 39 |
| ▪ Coronary Artery Bypass Graft mortality rate | Number of deaths per 100 discharges with procedure code for CABG in any field.<br><br>Age 40 years and older. | All non-maternal/non-neonatal discharges with procedure code for CABG (see below) in any field. Age 40 years and older.<br><br>Exclude transfers to other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates) | 40 |
| ▪ Craniotomy mortality rate | Number of deaths per 100 discharges with DRG code for craniotomy (DRG 001, Craniotomy Age > 17, Except for Trauma). | All non-maternal/non-neonatal discharges with DRG code for craniotomy (DRG 001 Craniotomy Age > 17, Except for Trauma).<br><br>Exclude transfers to other institution. | 41 |
| ▪ Esophageal resection mortality rate | Number of deaths per 100 esophageal cancer discharges with procedure code for esophageal resection for in any field. | All non-maternal/non-neonatal esophageal cancer discharges with procedure code for esophageal resection (see below) in any field.<br><br>Exclude transfers to other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | 42 |
| ▪ Hip replacement mortality rate | Number of deaths per 100 discharges with procedure code for partial or full hip replacement in any field. | All non-maternal/non-neonatal discharges with procedure code for partial or full hip replacement. (see below) in any field. | 43 |

| Indicator | Numerator | Denominator | # |
|---|---|---|---|
| | | Include only discharges with uncomplicated cases: diagnosis or procedure codes for osteoarthrosis of hip. (see below) in any field.<br><br>Exclude transfers to other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | |
| ▪ Pancreatic resection mortality rate | Number of deaths per 100 pancreatic cancer discharges with procedure code for pancreatic resection in any field. | All non-maternal/non-neonatal pancreatic cancer discharges with procedure code for pancreatic resection (see below) in any field.<br><br>Exclude transfers to other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium) and MDC 15 (newborns and other neonates). | 44 |
| ▪ Pediatric heart surgery mortality rate | Number of deaths per 100 discharges with procedure code for pediatric heart surgery in any field. | All non-maternal/non-neonatal discharges less than 18 with procedure code for specified heart surgery (see below) in any field, or for any heart surgery and a diagnosis code for hypoplastic left heart syndrome (see below) in any field.<br><br>See NOTE for additional exclusions<br>Exclude transfers to other institution.<br>Exclude MDC 14 (pregnancy, childbirth, and puerperium). | 45 |

NOTE (Exclusions for Pediatric Heart Surgery): 1) patients who underwent PDA ligation as a single cardiac procedure (diagnosis code 747.0 and procedure code 38.85); 2) patients with prosthetic closures of atrial septal defects (procedure codes 35.51-52, 35.71) or ventricular septal defects (codes 35.53, 35.72) or atrial septal enlargement (35.41) without concomitant use of cardiopulmonary bypass (code 39.61); 3) patients with PDA closure as single cardiac procedure (procedure code 38.85) with concomitant cardiac catheterization (codes 37.21, 37.22, 37.23, 88.42, 88.43); 4) patients with occlusion of thoracic vessel (procedure code 38.85) without congenital heart defect (diagnosis codes 745.0-747.9).

| Indicator | Coding Source |
|---|---|
| ▪ AAA repair volume and mortality rate | CCS Procedure Category #52<br>Manheim LM, J Vasc Surg 1998; 28:48-58. |
| ▪ Acute myocardial infarction mortality rate | CCS Diagnosis Category #100 |
| ▪ Adult asthma admission rate | HCUP-3 QI (AHRQ Pub. No. 98-0035, page 85) |
| ▪ Angina admission rate | Billings et al, "Analysis of Variation in Hospital Admission Rates Associated with Area Income in New York City," (March, 1992) |
| ▪ Bacterial pneumonia admission rate | Billings et al, "Analysis of Variation in Hospital Admission Rates Associated with Area Income in New York City," (March, 1992) |
| ▪ Bi-lateral catheterization rate | Medicare Quality of Care Report of Surveillance Measures (http://www.hcfa.gov/quality/3d.htm#2 ). |
| ▪ CABG volume, utilization rate, and mortality rate | CCS Procedure Category #44 (except 36.2, 36.3x) |
| ▪ Carotid Endarterectomy volume and utilization rate | CCS Procedure Category #51 |
| ▪ Cesarean section delivery rate | HCUP-3 QI (AHRQ Pub. No. 98-0035, page 82) |
| ▪ Chronic Obstructive Pulmonary Disease admission rate | CCS Diagnosis Category #127 |
| ▪ Congestive Heart Failure admission rate and mortality rate | CCS Diagnosis Category #108 |
| ▪ Craniotomy mortality rate | DRG 001, "CRANIOTOMY AGE >17 EXCEPT FOR TRAUMA" |
| ▪ Dehydration admission rate | Billings et al, "Analysis of Variation in Hospital Admission Rates Associated with Area Income in New York City," (March, 1992) |
| ▪ Diabetes – long-term complications admission rate | HCUP-3 QI (AHRQ Pub. No. 98-0035, page 86) |
| ▪ Diabetes – short-term complications admission rate | HCUP-3 QI (AHRQ Pub. No. 98-0035, page 86) |
| ▪ Diabetes – uncontrolled without complications admission rate | Healthy People 2010 |
| ▪ Esophageal resection volume and mortality rate | Patti MG, Corvera CU, Glasgow RE, Way LW. A hospital's annual rate of esophagectomy influences the operative mortality rate. J Gastrointest Surg. 1998; 2:186-192.<br><br>Begg CB, Cramer LD, Hoskins WJ, Brennan MF. Impact of hospital volume on operative mortality for major cancer surgery. JAMA. 1998; 280:1747-1751 |
| ▪ GI Hemorrhage mortality rate | HCUP-3 QI (AHRQ Pub. No. 98-0035, page 75) |
| ▪ Hip fracture mortality rate | CCS Diagnosis Category #226 |
| ▪ Hip replacement mortality rate | HCUP-3 QI (AHRQ Pub. No. 98-0035, page 73) |
| ▪ Hypertension admission rate | Billings et al, "Analysis of Variation in Hospital Admission Rates Associated with Area Income in New York City," (March, 1992) |
| ▪ Hysterectomy rate | CCS Procedure Category #124<br>HCUP-3 QI (AHRQ Pub. No. 98-0035, page 83) |
| ▪ Incidental appendectomy rate (Utilization #5) | HCUP-3 QI (AHRQ Pub. No. 98-0035, page 82) |

| Indicator | Coding Source |
|---|---|
| ▪ Laminectomy and/or spinal fusion rate | HCUP-3 QI (AHRQ Pub. No. 98-0035, page 83) |
| ▪ Laparoscopic cholecystectomy rate | HCUP-3 QI (AHRQ Pub. No. 98-0035, page 84) |
| ▪ Low birthweight rate | HCUP-3 QI (AHRQ Pub. No. 98-0035, page 84) |
| ▪ Lower extremity amputation in diabetics admission rate | Healthy People 2010 |
| ▪ Pancreatic resection volume and mortality rate | Begg CB, Cramer LD, Hoskins WJ, Brennan MF. Impact of hospital volume on operative mortality for major cancer surgery. JAMA. 1998; 280:1747-1751.<br>Glasgow RE, Mulvihill SJ. Hospital volume influences outcome in patients undergoing pancreatic resection for cancer. West J Med. 1996; 165:294-300.<br>Gordon TA, Bowman HM, Tielsch JM, Bass EB, Burleyson GP, Cameron JL. Statewide regionalization of pancreaticoduodenectomy and its effect on in-hospital mortality. Ann Surg. 1998; 228:71-78.<br>Imperato PJ, Nenner RP, Starr HA, Will TO, Rosenberg CR, Dearie MB. The effects of regionalization on clinical outcomes for a high-risk surgical procedure. Am J Med Qual. 1996; 11: 193-197.<br>Lieberman MD, Kilburn H, Lindsey M, Brennan MF.Relation of perioperative deaths to hospital volume among patients undergoing pancreatic resection for malignancy. *Ann Surg.* 1995;222:638-645. |
| ▪ Pediatric asthma admission rate | HCUP-3 QI (AHRQ Pub. No. 98-0035, page 85) |
| ▪ Pediatric gastroenteritis admission rate | Billings et al, "Analysis of Variation in Hospital Admission Rates Associated with Area Income in New York City," (March, 1992) |
| ▪ Pediatric heart surgery volume and mortality rate | Jenkins KJ, et al. Pediatrics 1995; 95: 323-330. |
| ▪ Perforated appendix admission rate | HCUP-3 QI (AHRQ Pub. No. 98-0035, page 87) |
| ▪ Pneumonia mortality rate | CCS Diagnosis Category #122 |
| ▪ PTCA volume and utilization rate | CCS Procedure Category #45 |
| ▪ Stroke mortality rate | HCUP-3 QI (AHRQ Pub. No. 98-0035, page 85) |
| ▪ Urinary infection admission rate | Billings et al, "Analysis of Variation in Hospital Admission Rates Associated with Area Income in New York City," (March, 1992) |
| ▪ Vaginal birth after cesarean section (VBAC) rate | HCUP-3 QI (AHRQ Pub. No. 98-0035, page 82) |

## Volume

***AAA repair volume(QI #1)***

**AAA repair**

ICD-9-CM procedure codes:
3834 AORTA RESECTION & ANAST
3844 RESECT ABDM AORTA W REPL
3864 EXCISION OF AORTA

*Include only:* **AAA**

ICD-9-CM diagnosis codes:
4413 RUPT ABD AORTIC ANEURYSM
4414 ABDOM AORTIC ANEURYSM

***Carotid endarterectomy volume(QI #2)***

**Carotid endarterectomy**

ICD-9-CM procedure codes:
3812 HEAD & NECK ENDARTER NEC

***CABG volume(QI #3)***

**CABG**

ICD-9-CM procedure codes:
3610 AORTOCORONARY BYPASS NOS
3611 AORTOCOR BYPAS-1 COR ART
3612 AORTOCOR BYPAS-2 COR ART
3613 AORTOCOR BYPAS-3 COR ART
3614 AORTCOR BYPAS-4+ COR ART
3615 INT MAM-COR ART BYPASS
3616 INT MAM-COR ART BYPASS
3617 ABD-CORON ART BYPASS OCT96-
3619 HRT REVAS BYPS ANAS NEC

***Esophageal resection volume(QI #4)***

**Esophageal resection**

ICD-9-CM procedure codes:
4240 ESOPHAGECTOMY NOS
4241 PARTIAL ESOPHAGECTOMY
4242 TOTAL ESOPHAGECTOMY

*Include only:* **Esophageal cancer:**

ICD-9-CM diagnosis codes:
1500 MAL NEO CERVICAL ESOPHAG
1501 MAL NEO THORACIC ESOPHAG
1502 MAL NEO ABDOMIN ESOPHAG
1503 MAL NEO UPPER 3RD ESOPH
1504 MAL NEO MIDDLE 3RD ESOPH
1505 MAL NEO LOWER 3RD ESOPH
1508 MAL NEO ESOPHAGUS NEC
1509 MAL NEO ESOPHAGUS NOS

***Pancreatic resection volume(QI #5)***

**Pancreatic resection**

ICD-9-CM procedure codes:
526 TOTAL PANCREATECTOMY
527 RAD PANCREATICODUODENECT

*Include only:* **Pancreatic cancer**

ICD-9-CM diagnosis codes:
1520 MALIGNANT NEOPL DUODENUM
1561 MAL NEO EXTRAHEPAT DUCTS
1562 MAL NEO AMPULLA OF VATER
1570 MAL NEO PANCREAS HEAD
1571 MAL NEO PANCREAS BODY
1572 MAL NEO PANCREAS TAIL
1573 MAL NEO PANCREATIC DUCT
1574 MAL NEO ISLET LANGERHANS
1578 MALIG NEO PANCREAS NEC
1579 MALIG NEO PANCREAS NOS

***Pediatric heart surgery volume(QI #6)***

**Pediatric heart surgery (#1)**

ICD-9-CM procedure codes:
3500 CLOSED VALVOTOMY NOS
3501 CLOSED AORTIC VALVOTOMY
3502 CLOSED MITRAL VALVOTOMY
3503 CLOSED PULMON VALVOTOMY
3504 CLOSED TRICUSP VALVOTOMY
3510 OPEN VALVULOPLASTY NOS
3511 OPN AORTIC VALVULOPLASTY
3512 OPN MITRAL VALVULOPLASTY
3513 OPN PULMON VALVULOPLASTY
3514 OPN TRICUSP VALVULOPLASTY
3520 REPLACE HEART VALVE NOS
3521 REPLACE AORT VALV-TISSUE
3522 REPLACE AORTIC VALVE NEC
3523 REPLACE MITR VALV-TISSUE
3524 REPLACE MITRAL VALVE NEC
3525 REPLACE PULM VALV-TISSUE
3526 REPLACE PULMON VALVE NEC
3527 REPLACE TRIC VALV-TISSUE
3528 REPLACE TRICUSP VALV NEC
3531 PAPILLARY MUSCLE OPS
3532 CHORDAE TENDINEAE OPS
3533 ANNULOPLASTY
3534 INFUNDIBULECTOMY
3535 TRABECUL CARNEAE CORD OP
3539 TISS ADJ TO VALV OPS NEC
3542 CREATE SEPTAL DEFECT
3550 PROSTH REP HRT SEPTA NOS
3554 PROS REP ENDOCAR CUSHION
3560 GRFT REPAIR HRT SEPT NOS
3561 GRAFT REPAIR ATRIAL DEF
3562 GRAFT REPAIR VENTRIC DEF
3563 GRFT REP ENDOCAR CUSHION
3570 HEART SEPTA REPAIR NOS
3573 ENDOCAR CUSHION REP NEC
3581 TOT REPAIR TETRAL FALLOT
3582 TOTAL REPAIR OF TAPVC
3583 TOT REP TRUNCUS ARTERIOS
3584 TOT COR TRANSPOS GRT VES OCT88-
3591 INTERAT VEN RETRN TRANSP
3592 CONDUIT RT VENT-PUL ART
3593 CONDUIT LEFT VENTR-AORTA
3594 CONDUIT ARTIUM-PULM ART
3595 HEART REPAIR REVISION
3598 OTHER HEART SEPTA OPS
3599 OTHER OP ON HRT VALVES
3835 THOR VESSEL RESECT/ANAST
3845 RESECT THORAC VES W REPL
390 SYSTEMIC-PULM ART SHUNT
3921 CAVAL-PULMON ART ANASTOM
3959 REPAIR OF VESSEL NEC

**Pediatric heart surgery (#1A)**

ICD-9-CM procedure codes:
3541 ENLARGE EXISTING SEP DEF#
3551 PROS REP ATRIAL DEF-OPN#
3552 PROS REPAIR ATRIA DEF-CL#
3553 PROST REPAIR VENTRIC DEF#
3571 ATRIA SEPTA DEF REP NEC#
3572 VENTR SEPTA DEF REP NEC#

*Include only:* Cardiopulmonary Bypass

ICD-9-CM procedure code:
3961 EXTRACORPOREAL CIRCULAT

**Pediatric heart surgery (#1B)**

ICD-9-CM procedure code:
3885 OCCLUDE THORACIC VES NEC*

*Include only*: **Congenital heart defect**

ICD-9-CM diagnosis codes:
7450 COMMON TRUNCUS
74510 COMPL TRANSPOS GREAT VES
74511 DOUBLE OUTLET RT VENTRIC
74512 CORRECT TRANSPOS GRT VES
74519 TRANSPOS GREAT VESS NEC
7452 TETRALOGY OF FALLOT
7453 COMMON VENTRICLE
7454 VENTRICULAR SEPT DEFECT
7455 SECUNDUM ATRIAL SEPT DEF
74560 ENDOCARD CUSHION DEF NOS
74561 OSTIUM PRIMUM DEFECT
74569 ENDOCARD CUSHION DEF NEC
7457 COR BILOCULARE
7458 SEPTAL CLOSURE ANOM NEC
7459 SEPTAL CLOSURE ANOM NOS
74600 PULMONARY VALVE ANOM NOS
74601 CONG PULMON VALV ATRESIA
74602 CONG PULMON VALVE STENOS
74609 PULMONARY VALVE ANOM NEC
7461 CONG TRICUSP ATRES/STEN
7462 EBSTEIN'S ANOMALY
7463 CONG AORTA VALV STENOSIS
7464 CONG AORTA VALV INSUFFIC
7465 CONGEN MITRAL STENOSIS
7466 CONG MITRAL INSUFFICIENC
7467 HYPOPLAS LEFT HEART SYND
74681 CONG SUBAORTIC STENOSIS
74682 COR TRIATRIATUM
74683 INFUNDIB PULMON STENOSIS
74684 OBSTRUCT HEART ANOM NEC
74685 CORONARY ARTERY ANOMALY
74686 CONGENITAL HEART BLOCK
74687 MALPOSITION OF HEART
74689 CONG HEART ANOMALY NEC
7469 CONG HEART ANOMALY NOS
7470 PATENT DUCTUS ARTERIOSUS
74710 COARCTATION OF AORTA
74711 INTERRUPT OF AORTIC ARCH
74720 CONG ANOM OF AORTA NOS
74721 ANOMALIES OF AORTIC ARCH
74722 AORTIC ATRESIA/STENOSIS
74729 CONG ANOM OF AORTA NEC
7473 PULMONARY ARTERY ANOM
74740 GREAT VEIN ANOMALY NOS
74741 TOT ANOM PULM VEN CONNEC
74742 PART ANOM PULM VEN CONN
74749 GREAT VEIN ANOMALY NEC
7475 UMBILICAL ARTERY ABSENCE
74760 UNSP PRPHERL VASC ANOMAL
74761 GSTRONTEST VESL ANOMALY
74762 RENAL VESSEL ANOMALY
74763 UPR LIMB VESSEL ANOMALY
74764 LWR LIMB VESSEL ANOMALY
74769 OTH SPCF PRPH VSCL ANOML
74781 CEREBROVASCULAR ANOMALY
74782 SPINAL VESSEL ANOMALY
74789 CIRCULATORY ANOMALY NEC
7479 CIRCULATORY ANOMALY NOS

*Or*

*Exclude* (if single procedure): **PDA ligation**

ICD-9-CM diagnosis code:
7470 PATENT DUCTUS ARTERIOSUS

*Or*

*Exclude* (if single procedure): **Cardiac Catheterization**

ICD-9-CM procedure codes:
3721 RT HEART CARDIAC CATH
3722 LEFT HEART CARDIAC CATH
3723 RT/LEFT HEART CARD CATH
8842 CONTRAST AORTOGRAM
8843 CONTR PULMON ARTERIOGRAM

**Pediatric heart surge ry (#2)**

ICD-9-CM procedure codes:
3500 CLOSED VALVOTOMY NOS
3501 CLOSED AORTIC VALVOTOMY
3502 CLOSED MITRAL VALVOTOMY
3503 CLOSED PULMON VALVOTOMY
3504 CLOSED TRICUSP VALVOTOMY
3510 OPEN VALVULOPLASTY NOS
3511 OPN AORTIC VALVULOPLASTY
3512 OPN MIT RAL VALVULOPLASTY
3513 OPN PULMON VALVULOPLASTY
3514 OPN TRICUS VALVULOPLASTY
3520 REPLACE HEART VALVE NOS
3521 REPLACE AORT VALV-TISSUE
3522 REPLACE AORTIC VALVE NEC
3523 REPLACE MITR VALV-TISSUE
3524 REPLACE MITRAL VALVE NEC
3525 REPLACE PULM VALV-TISSUE
3526 REPLACE PULMON VALVE NEC
3527 REPLACE TRIC VALV-TISSUE
3528 REPLACE TRICUSP VALV NEC
3531 PAPILLARY MUSCLE OPS
3532 CHORDAE TENDINEAE OPS
3533 ANNULOPLASTY
3534 INFUNDIBULECTOMY
3535 TRABECUL CARNEAE CORD OP
3539 TISS ADJ TO VALV OPSNEC
3542 CREATE SEPTAL DEFECT
3550 PROSTH REP HRT SEPTA NOS
3554 PROS REP ENDOCAR CUSHION
3560 GRFT REPAIR HRT SEPT NOS
3561 GRAFT REPAIR ATRIAL DEF
3562 GRAFT REPAIR VENTRIC DEF
3563 GRFT REP ENDOCAR CUSHION
3570 HEART SEPTA REPAIR NOS
3573 ENDOCAR CUSHION REP NEC
3581 TOT REPAIR TETRAL FALLOT
3582 TOTAL REPAIR OF TAPVC
3583 TOT REP TRUNCUS ARTERIOS
3584 TOT COR TRANSPOS GRT VES
3591 INTERAT VEN RETRN TRANSP
3592 CONDUIT RT VENT-PUL ART
3593 CONDUIT LEFT VENTR-AORTA
3594 CONDUIT ARTIUM-PULM ART
3595 HEART REPAIR REVISION
3596 PERC HEART VALVULOPLASTY
3598 OTHER HEART SEPTA OPS
3599 OTHER HEART VALVE OPS
3601 PTCA-1 VES/ATH W/O AGENT
3602 PTCA-1 VES/ATH W AGENT
3603 OPEN CORONRY ANGIOPLASTY
3604 INTRCORONRY THROMB INFUS
3605 PTCA-MULTIPLE VESSEL/ATH
3606 INSERT OF COR ART STENT OCT95-
3609 REM OF COR ART OBSTR NEC
3610 AORTOCORONARY BYPASS NOS
3611 AORTOCOR BYPAS-1 COR ART
3612 AORTOCOR BYPAS-2 COR ART

3613 AORTOCOR BYPAS-3 COR ART
3614 AORTCOR BYPAS-4+ COR ART
3615 INT MAM-COR ART BYPASS
3616 INT MAM-COR ART BYPASS
3617 ABD-CORON ARTERY BYPASS OCT96-
3619 HRT REVAS BYPS ANAS NEC
362 ARTERIAL IMPLANT REVASC
363 HEART REVASCULARIZAT NEC –OCT98
3631 OPEN CHEST TRANS REVASC OCT98-
3632 OTH TRANSMYO REVASCULAR OCT98-
3639 OTH HEART REVASCULAR OCT98-
3691 CORON VESS ANEURYSM REP
3699 HEART VESSEL OP NEC
370 PERICARDIOCENTESIS
3710 INCISION OF HEART NOS
3711 CARDIOTOMY
3712 PERICARDIOTOMY
3721 RT HEART CARDIAC CATH
3722 LEFT HEART CARDIAC CATH
3723 RT/LEFT HEART CARD CATH
3724 PERICARDIAL BIOPSY
3725 CARDIAC BIOPSY
3726 CARDIAC ELECTROPHY STIM
3727 CARDIAC MAPPING
3729 HRT/PERICAR DX PROC NEC
3731 PERICARDIECTOMY
3732 HEART ANEURYSM EXCISION
3733 EXC/DEST OTH HRT LESION
3734 CATH ABLATION LES HEART
3735 PARTIAL VENTRICULECTOMY OCT97-
374 HEART & PERICARD REPAIR
375 HEART TRANSPLANTATION
3761 PULSATION BALLOON IMPLAN
3762 IMPLANT HRT ASST SYS NEC
3763 REPLACE HRT ASSIST SYST
3764 REMOVE HEART ASSIST SYS
3765 IMP EXT PUL HRT ASST SYS OCT95-
3766 IMP IMP PUL HRT ASST SYS OCT95-
3767 IMP CARDIOMYOSTIMUL SYS OCT98-
3770 INT INSERT PACEMAK LEAD
3771 INT INSERT LEAD IN VENT
3772 INT INSER LEAD ATRI-VENT
3773 INT INSER LEAD IN ATRIUM
3774 INT OR REPL LEAD EPICAR
3775 REVISION OF LEAD
3776 REPL TV ATRI-VENT LEAD
3777 REMOVAL OF LEAD W/O REPL
3778 INSER TEMP PACEMAKER SYS
3779 REVIS OR RELOCATE POCKET
3780 INT OR REPL PERM PACEMKR
3781 INT INSERT 1-CHAM, NON
3782 INT INSERT 1-CHAM, RATE
3783 INT INSERT DUAL-CHAM DEV
3785 REPL PACEM W 1-CHAM, NON
3786 REPL PACEM 1-CHAM, RATE
3787 REPL PACEM W DUAL-CHAM
3789 REVISE OR REMOVE PACEMAK
3791 OPN CHEST CARDIAC MASSAG
3792 INJECTION INTO HEART
3793 INJECTION INTO PERICARD
3794 IMPLT/REPL CARDDEFIB TOT
3795 IMPLT CARDIODEFIB LEADS
3796 IMPLT CARDIODEFIB GENATR
3797 REPL CARDIODEFIB LEADS
3798 REPL CARDIODEFIB GENRATR
3799 OTHER HEART/PERICARD OPS

*Include only:* **Hypoplastic left heart syndrome:**

ICD-9-CM diagnosis code:
7467 HYPOPLAS LEFT HEART SYND

***PTCA Volume(QI #7)***

**PTCA**

ICD-9-CM procedure codes:
3601 PTCA-1 VESSEL W/O AGENT
3602 PTCA-1 VESSEL WITH AGNT
3605 PTCA-MULTIPLE VESSEL
3606 INSERT OF COR ART STENT OCT95-

## Utilization – Provider level

***Cesarean section delivery(QI #8)***

**Cesarean section deliveries (Outcome of Interest):**

DRGs:
370 Cesarean section w/CC
371 Cesarean section w/o CC

**All deliveries (Population at Risk):**

DRGs:
370 CESAREAN SECTION W CC
371 CESAREAN SECTION W/O CC
372 VAGINAL DELIVERY W COMPL
373 VAG DELIVERY W/O COMPL
374 VAG DELIV W STERIL OR DC
375 VAG DELIV W OTH OR PROC

***Incidental appendectomy among elderly(QI #9)***

**Incidental appendectomy (Outcome of Interest):**

ICD-9-CM procedure codes:
471 INCIDENTAL APPENDECTOMY -OCT96
4711 LAP INCID APPENDECTOMY OCT96-
4719 OTHER INCID APPENDECTOMY OCT96-

**Intra-abdominal procedure (Population at Risk):**

DRGs:
146 RECTAL RESECTION W CC
147 RECTAL RESECTION W/O CC
148 MAJ BOWEL PROC W CC
149 MAJ BOWEL PROC W/0 CC
150 PERITONEAL ADHES W CC
151 PERITONEAL ADHES W/O CC
152 MIN BOWEL PROC W CC
153 MIN BOWEL PROC W/O CC
154 UGI PROC AGE >17 W CC
155 UGI PROC AGE >17 W/O CC
170 OTH GI OR PROC W CC
171 OTH GI OR PROC W/O CC
191 PANC LVR SHNT PRC W CC
192 PANC LVR SHNT PRC W/O CC
193 BILIARY PROC W/ CC
194 BILIARY PROC W/O CC
195 CHOLE W/ CDE W/ CC
196 CHOLE W/ CDE W/O CC
197 CHOLE W/ CC
198 CHOLE W/O CC
201 OTH BILIARY/PANC PROC
354 UTER PROC MALIG W/ CC
355 UTER PROC MALIG W/O CC
356 FEMALE REPROD RECONSTR
357 UTER PROC OVARIAN MALIG
358 UTER PROC NONMALIG W/ CC
359 UTER PROC NONMALI W/O CC
365 OTH FEMAL REPROD PROC

***Bilateral cardiac catheterization(QI #10)***

**Bilateral catheterization (Outcome of interest)**

ICD-9-CM procedure code:
3723 RT/LEFT HEART CARD CATH

**Catheterization (Population at risk)**

ICD-9-CM procedure codes:

| | |
|---|---|
| 3722 | LEFT HEART CARDIAC CATH |
| 3723 | RT/LEFT HEART CARD CATH |

*Include only:* **Coronary artery disease**

ICD-9-CM diagnosis codes:

| | |
|---|---|
| 41000 | AMI ANTEROLATERAL,UNSPEC |
| 41001 | AMI ANTEROLATERAL, INIT |
| 41002 | AMI ANTEROLATERAL,SUBSEQ |
| 41010 | AMI ANTERIOR WALL,UNSPEC |
| 41011 | AMI ANTERIOR WALL, INIT |
| 41012 | AMI ANTERIOR WALL,SUBSEQ |
| 41020 | AMI INFEROLATERAL,UNSPEC |
| 41021 | AMI INFEROLATERAL, INIT |
| 41022 | AMI INFEROLATERAL,SUBSEQ |
| 41030 | AMI INFEROPOST, UNSPEC |
| 41031 | AMI INFEROPOST, INITIAL |
| 41032 | AMI INFEROPOST, SUBSEQ |
| 41040 | AMI INFERIOR WALL,UNSPEC |
| 41041 | AMI INFERIOR WALL, INIT |
| 41042 | AMI INFERIOR WALL,SUBSEQ |
| 41050 | AMI LATERAL NEC, UNSPEC |
| 41051 | AMI LATERAL NEC, INITIAL |
| 41052 | AMI LATERAL NEC, SUBSEQ |
| 41060 | TRUE POST INFARCT,UNSPEC |
| 41061 | TRUE POST INFARCT, INIT |
| 41062 | TRUE POST INFARCT,SUBSEQ |
| 41070 | SUBENDO INFARCT, UNSPEC |
| 41071 | SUBENDO INFARCT, INITIAL |
| 41072 | SUBENDO INFARCT, SUBSEQ |
| 41080 | AMI NEC, UNSPECIFIED |
| 41081 | AMI NEC, INITIAL |
| 41082 | AMI NEC, SUBSEQUENT |
| 41090 | AMI NOS, UNSPECIFIED |
| 41091 | AMI NOS, INITIAL |
| 41092 | AMI NOS, SUBSEQUENT |
| 4110 | POST MI SYNDROME |
| 4111 | INTERMED CORONARY SYND |
| 41181 | CORONARY OCCL SN W/O MI |
| 41189 | AC ISCHEMIC HRT DIS NEC |
| 412 | OLD MYOCARDIAL INFARCT |
| 4130 | ANGINA DECUBITUS |
| 4131 | PRINZMETAL ANGINA |
| 4139 | ANGINA PECTORIS NEC/NOS |
| 4140 | CORONARY ATHEROSCLEROSIS –OCT94 |
| 41400 | COR ATH UNSP VSL NTV/GFT OCT94- |
| 41401 | CRNRY ATHRSCL NATVE VSSL OCT94- |
| 41402 | CRN ATH ATLG VN BPS GRFT OCT94- |
| 41403 | CRN ATH NONATLG BLG GRFT OCT94- |
| 41404 | COR ATH ARTRY BYPAS GRFT OCT96- |
| 41405 | COR ATH BYPASS GRAFT NOS OCT96- |
| 41410 | ANEURYSM, HEART (WALL) |
| 41411 | CORONARY VESSEL ANEURYSM |
| 41419 | ANEURYSM OF HEART NEC |
| 4148 | CHR ISCHEMIC HRT DIS NEC |
| 4149 | CHR ISCHEMIC HRT DIS NOS |

***Excludes:*** **Possible indications for right cath**

ICD-9-CM diagnosis codes:

| | |
|---|---|
| 3910 | ACUTE RHEUMATIC PERICARD |
| 3911 | ACUTE RHEUMATIC ENDOCARD |
| 3912 | AC RHEUMATIC MYOCARDITIS |
| 3918 | AC RHEUMAT HRT DIS NEC |
| 3919 | AC RHEUMAT HRT DIS NOS |
| 3920 | RHEUM CHOREA W HRT INVOL |
| 3929 | RHEUMATIC CHOREA NOS |
| 393 | CHR RHEUMATIC PERICARD |
| 3940 | MITRAL STENOSIS |
| 3941 | RHEUMATIC MITRAL INSUFF |
| 3942 | MITRAL STENOSIS W INSUFF |
| 3949 | MITRAL VALVE DIS NEC/NOS |
| 3960 | MIT RAL/AORTIC STENOSIS |
| 3961 | MITRAL STENOS/AORT INSUF |
| 3962 | MITRAL INSUF/AORT STENOS |
| 3963 | MITRAL/AORTIC VAL INSUFF |
| 3968 | MITR/AORTIC MULT INVOLV |
| 3969 | MITRAL/AORTIC V DIS NOS |
| 3970 | TRICUSPID VALVE DISEASE |
| 3971 | RHEUM PULMON VALVE DIS |
| 3979 | RHEUM ENDOCARDITIS NOS |
| 3980 | RHEUMATIC MYOCARDITIS |
| 39890 | RHEUMATIC HEART DIS NOS |
| 39891 | RHEUMATIC HEART FAILURE |
| 39899 | RHEUMATIC HEART DIS NEC |
| 40200 | MAL HYPERTEN HRT DIS NOS |
| 40201 | MAL HYPERT HRT DIS W CHF |
| 40210 | BEN HYPERTEN HRT DIS NOS |
| 40211 | BENIGN HYP HRT DIS W CHF |
| 40290 | HYPERTENSIVE HRT DIS NOS |
| 40291 | HYPERTEN HEART DIS W CHF |
| 4150 | ACUTE COR PULMONALE |
| 4151 | PULMON EMBOLISM/INFARCT –OCT95 |
| 41511 | IATROGEN PULM EMB/INFARC OCT95- |
| 41519 | PULM EMBOL/INFARCT NEC OCT95- |
| 4160 | PRIM PULM HYPERTENSION |
| 4161 | KYPHOSCOLIOTIC HEART DIS |
| 4168 | CHR PULMON HEART DIS NEC |
| 4169 | CHR PULMON HEART DIS NOS |
| 4170 | ARTERIOVEN FISTU PUL VES |
| 4171 | PULMON ARTERY ANEURYSM |
| 4178 | PULMON CIRCULAT DIS NEC |
| 4179 | PULMON CIRCULAT DIS NOS |
| 4242 | NONRHEUM TRICUSP VAL DIS |
| 4243 | PULMONARY VALVE DISORDER |
| 7450 | COMMON TRUNCUS |
| 74510 | COMPL TRANSPOS GREAT VES |
| 74511 | DOUBLE OUTLET RT VENTRIC |
| 74512 | CORRECT TRANSPOS GRT VES |
| 74519 | TRANSPOS GREAT VESS NEC |
| 7452 | TETRALOGY OF FALLOT |
| 7453 | COMMON VENTRICLE |
| 7454 | VENTRICULAR SEPT DEFECT |
| 7455 | SECUNDUM ATRIAL SEPT DEF |
| 74560 | ENDOCARD CUSHION DEF NOS |
| 74561 | OSTIUM PRIMUM DEFECT |
| 74569 | ENDOCARD CUSHION DEF NEC |
| 7457 | COR BILOCULARE |
| 7458 | SEPTAL CLOSURE ANOM NEC |
| 7459 | SEPTAL CLOSURE ANOM NOS |
| 74600 | PULMONARY VALVE ANOM NOS |
| 74601 | CONG PULMON VALV ATRESIA |
| 74602 | CONG PULMON VALVE STENOS |
| 74609 | PULMONARY VALVE ANOM NEC |
| 7461 | CONG TRICUSP ATRES/STEN |
| 7462 | EBSTEIN'S ANOMALY |
| 7463 | CONG AORTA VALV STENOSIS |
| 7464 | CONG AORTA VALV INSUFFIC |
| 7465 | CONGEN MITRAL STENOSIS |
| 7466 | CONG MITRAL INSUFFICIENC |
| 7467 | HYPOPLAS LEFT HEART SYND |
| 74681 | CONG SUBAORTIC STENOSIS |
| 74682 | COR TRIATRIATUM |
| 74683 | INFUNDIB PULMON STENOSIS |
| 74684 | OBSTRUCT HEART ANOM NEC |
| 74685 | CORONARY ARTERY ANOMALY |
| 74686 | CONGENITAL HEART BLOCK |
| 74687 | MALPOSITION OF HEART |
| 74689 | CONG HEART ANOMALY NEC |
| 7469 | CONG HEART ANOMALY NOS |
| 7470 | PATENT DUCTUS ARTERIOSUS |
| 74710 | COARCTATION OF AORTA |
| 74711 | INTERRUPT OF AORTIC ARCH |
| 74720 | CONG ANOM OF AORTA NOS |
| 74721 | ANOMALIES OF AORTIC ARCH |
| 74722 | AORTIC ATRESIA/STENOSIS |
| 74729 | CONG ANOM OF AORTA NEC |

7473 PULMONARY ARTERY ANOM
74740 GREAT VEIN ANOMALY NOS
74741 TOT ANOM PULM VEN CONNEC
74742 PART ANOM PULM VEN CONN
74749 GREAT VEIN ANOMALY NEC
7475 UMBILICAL ARTERY ABSENCE
74760 UNSP PRPHERL VASC ANOMAL
74761 GSTRONTEST VESL ANOMALY
74762 RENAL VESSEL ANOMALY
74763 UPR LIMB VESSEL ANOMALY
74764 LWR LIMB VESSEL ANOMALY
74769 OTH SPCF PRPH VSCL ANOML
74781 CEREBROVASCULAR ANOMALY
74782 SPINAL VESSEL ANOMALY
74789 CIRCULATORY ANOMALY NEC
7479 CIRCULATORY ANOMALY NOS

***Successful vaginal birth after cesarean section (VBAC) (QI #11)***

**Vaginal deliveries (Outcome of Interest):**

DRGs:
372 VAGINAL DELIVERY W/ CC
373 VAGINAL DELIVERY W/O CC
374 VAGINAL DELIVERY W/ STERILIZATION OR D&C
375 VAGINAL DELIVERY W/ OTHER O.R. PROCEDURE

**All deliveries (Population at Risk):**

DRGs:
370 CESAREAN SECTION W CC
371 CESAREAN SECTION W/O CC
372 VAGINAL DELIVERY W COMPL
373 VAG DELIVERY W/O COMPL
374 VAG DELIV W STERIL OR DC
375 VAG DELIV W OTH OR PROC

**Previous C -section (Population at Risk):**

ICD-9-CM diagnosis codes:
65420 PREV C-SECT NOS-UNSPEC
65421 PREV C-SECT NOS-DELIVER
65423 PREV C-SECT NOS-ANTEPART

***Laparoscopic cholecystectomy rate(QI #12)***

**Laparoscopic cholecystectomy (Outcome of Interest):**

ICD-9-CM Procedure Code:
5123 LAPAROSCOPIC CHOLE

**Cholecystectomy (Population at Risk):**

ICD-9-CM Procedure Codes:
5122 CHOLECYSTECTOMY
5123 LAPAROSCOPIC CHOLE

*Include Only:* **Uncomplicated cholecystitis and/or cholelithiasis:**

ICD-9-CM diagnosis Codes:
57400 CHOLELITH W/ AC CHOLECYS
57401 CHOLELITH W/ AC CHOLECYS
57410 CHOLELITH W CHOLECYS NEC
57411 CHOLELITH W CHOLECYS NEC
57420 CHOLELITH NOS W/O OBSTRUC
57421 CHOLELITH NOS W/ OBSTRUC
5750 ACUTE CHOLECYSTITIS
5751 CHOLECYSTITIS NEC -OCT96
57510 CHOLECYSTITIS NOS OCT96-
57511 CHRONIC CHOLECYSTITIS OCT96-
57512 ACUTE & CHR CHOLECYSTITIS OCT96-

## Utilization – Area level

***CABG rate(QI #13)***

**CABG**

ICD-9-CM procedure codes:
3610 AORTOCORONARY BYPASS NOS
3611 AORTOCOR BYPAS-1 COR ART
3612 AORTOCOR BYPAS-2 COR ART
3613 AORTOCOR BYPAS-3 COR ART
3614 AORTCOR BYPAS-4+ COR ART
3615 1 INT MAM-COR ART BYPASS
3616 2 INT MAM-COR ART BYPASS
3617 ABD-CORON ART BYPASS OCT96-
3619 HRT REVAS BYPS ANAS NEC

***Hysterectomy rate(QI #14)***

**Hysterectomy**

ICD-9-CM Procedure Codes:
683 SUBTOT ABD HYSTERECTOMY
684 TOTAL ABD HYSTERECTOMY
685 VAGINAL HYSTERECTOMY -OCT96
6851 LAP AST VAG HYSTERECTOMY OCT96-
6859 OTHER VAG HYSTERECTOMY OCT96-
686 RADICAL ABD HYSTERECTOMY
687 RADICAL VAG HYSTERECTOMY
689 OTHER/UNSP HYSTERECTOMY

*Exclude:* **Female genital cancer:**

ICD-9-CM Diagnosis Codes:
179 MALIG NEOPL UTERUS NOS
1800 MALIG NEO ENDOCERVIX
1801 MALIG NEO EXOCERVIX
1808 MALIG NEO CERVIX NEC
1809 MAL NEO CERVIX UTERI NOS
181 MALIGNANT NEOPL PLACENTA
1820 MALIG NEO CORPUS UTERI
1821 MAL NEO UTERINE ISTHMUS
1828 MAL NEO BODY UTERUS NEC
1830 MALIGN NEOPL OVARY
1832 MAL NEO FALLOPIAN TUBE
1833 MAL NEO BROAD LIGAMENT
1834 MALIG NEO PARAMETRIUM
1835 MAL NEO ROUND LIGAMENT
1838 MAL NEO ADNEXA NEC
1839 MAL NEO ADNEXA NOS
1840 MALIGN NEOPL VAGINA
1841 MAL NEO LABIA MAJORA
1842 MAL NEO LABIA MINORA
1843 MALIGN NEOPL CLITORIS
1844 MALIGN NEOPL VULVA NOS
1848 MAL NEO FEMALE GENIT NEC
1849 MAL NEO FEMALE GENIT NOS
2331 CA IN SITU CERVIX UTERI
2332 CA IN SITU UTERUS NEC
2333 CA IN SITU FEM GEN NEC
2360 UNCERT BEHAV NEO UTERUS
2361 UNC BEHAV NEO PLACENTA
2362 UNC BEHAV NEO OVARY
2363 UNC BEHAV NEO FEMALE NEC

*Exclude:* **Pelvic or lower abdominal trauma:**

ICD-9-CM Diagnosis Codes:
8674 UTERUS INJURY-CLOSED
8675 UTERUS INJURY-OPEN
8676 PELVIC ORGAN INJ NEC-CL
8677 PELVIC ORGAN INJ NEC-OPN
8678 PELVIC ORGAN INJ NOS-CL
8679 PELVIC ORGAN INJ NOS-OPN
86800 INTRA-ABDOM INJ NOS-CLOS
86803 PERITONEUM INJURY-CLOSED
86804 RETROPERITONEUM INJ-CL
86809 INTRA-ABDOM INJ NEC-CLOS
86810 INTRA-ABDOM INJ NOS-OPEN
86813 PERITONEUM INJURY-OPEN
86814 RETROPERITONEUM INJ-OPEN
86819 INTRA-ABDOM INJ NEC-OPEN
8690 INTERNAL INJ NOS-CLOSED
8691 INTERNAL INJURY NOS-OPEN
8796 OPEN WOUND OF TRUNK NEC
8797 OPEN WND TRUNK NEC-COMPL
8798 OPEN WOUND SITE NOS
8799 OPN WOUND SITE NOS-COMPL
9060 LT EFF OPN WND HEAD/TRNK
9081 LATE EFF INT INJ ABDOMEN
9082 LATE EFF INT INJURY NEC
9391 FOREIGN BODY UTERUS
9474 BURN OF VAGINA & UTERUS

***Laminectomy and/or spinal fusion rate(QI #15)***

**Laminectomy and/or spinal fusion:**

ICD-9-CM Procedure Codes:
0302 REOPEN LAMINECTOMY SITE
0309 SPINAL CANAL EXPLOR NEC
8050 EXC/DEST INTVRT DISC NOS
8051 EXCISION INTERVERT DISC
8059 OTH EXC/DEST INTVRT DISC
8100 SPINAL FUSION NOS
8101 ATLAS-AXIS FUSION
8102 OTH CERV FUSION, ANTER
8103 OTH CERV FUSION, POSTER
8104 DORSAL FUSION, ANTERIOR
8105 DORSAL FUSION, POSTERIOR
8106 LUMBAR FUSION, ANTERI OR
8107 LUMBAR FUSION, LATERAL
8108 LUMBAR FUSION, POSTERIOR
8109 REFUSION OF SPINE

***PTCA rate(QI #16)***

**PTCA**

ICD-9-CM procedure codes:
3601 PTCA-1 VESSEL W/O AGENT
3602 PTCA-1 VESSEL WITH AGNT
3605 PTCA-MULTIPLE VESSEL
3606 INSERT OF COR ART STENT

## ACSC Area rates

***Dehydration rate(QI #17)***

**Dehydration**

ICD-9-CM diagnosis code:
2765 HYPOVOLEMIA

***Bacterial pneumonia rate(QI #18)***

**Bacterial pneumonia**

ICD-9-CM diagnosis codes:
481 PNEUMOCOCCAL PNEUMONIA
4822 H.INFLUENZAE PNEUMONIA
48230 STREP PNEUMONIA UNSPEC
48231 GRP A STREP PNEUMONIA
48232 GRP B STREP PNEUMONIA
48239 OTH STREP PNEUMONIA
4829 BACTERIAL PNEUMONIA NOS
4830 MYCOPLASMA PNEUMONIA
4831 CHLAMYDIA PNEUMONIA OCT96-
4838 OTH SPEC ORG PNEUMONIA
485 BRONCOPNEUMONIA ORG NOS
486 PNEUMONIA, ORGANISM NOS

*Exclude:* **Sickle cell anemia, HB disease**

ICD-9-CM diagnosis codes:
28260 SICKLE-CELL ANEMIA NOS
28261 HB-S DISEASE W/O CRISIS
28262 HB-S DISEASE WITH CRISIS
28263 SICKLE-CELL/HB-C DISEASE
28269 SICKLE-CELL ANEMIA NEC

***Urinary infection rate(QI #19)***

**Urinary infection**

ICD-9-CM diagnosis codes:
59000 CHR PYELONEPHRITIS NOS
59001 CHR PYELONEPH W MED NECR
59010 AC PYELONEPHRITIS NOS
59011 AC PYELONEPHR W MED NECR
5902 RENAL/PERIRENAL ABSCESS
5903 PYELOURETERITIS CYSTICA
59080 PYELONEPHRITIS NOS
59081 PYELONEPHRIT IN OTH DIS
5909 INFECTION OF KIDNEY NOS
5950 AC CYSTITIS
5959 CYSTITIS NOS
5990 URIN TRACT INFECTION NOS

***Perforated appendix rate(QI #20)***

**Perforated appendix (outcome of interest)**

ICD-9-CM diagnosis codes:
5400 AC APPEND W PERITONITIS
5401 ABSCESS OF APPENDIX

**Appendicitis (population at risk)**

ICD-9-CM diagnosis codes**:**
5400 AC APPEND W PERITONITIS
5401 ABSCESS OF APPENDIX
5409 ACUTE APPENDICITIS NOS
541 APPENDICITIS NOS

***Angina (w/o procedure) rate(QI #21)***

**Angina**

ICD-9-CM diagnosis codes:
4111 INTERMED CORONARY SYND
41181 CORONARY OCCLSN W/O MI
41189 AC ISCHEMIC HRT DIS NEC
4130 ANGINA DECUBITUS
4131 PRINZMETAL ANGINA
4139 ANGINA PECTORIS NEC/NOS

*Exclude*: **All procedures**:
(*01.0-86.99)*

***Adult asthma rate(QI #22)***

**Adult asthma**

ICD-9-CM diagnosis codes:

49300 EXT ASTHMA W/O STAT ASTH
49301 EXT ASTHMA W STATUS ASTH
49302 EXT ASTHMA W STATUS ACEX OCT00-
49310 INT ASTHMA W/O STAT ASTH
49311 INT ASTHMA W STATUS ASTH
49312 INT ASTHMA W STATUS ACEX OCT00-
49320 CH OB ASTH W/O STAT ASTH
49321 CH OB ASTHMA W STAT ASTH
49322 CH OB ASTHMA W STAT ACEX OCT00-
49390 ASTHMA W/O STATUS ASTHM
49391 ASTHMA W STATUS ASTHMAT
49392 ASTHMA W STATUS AC EXAC OCT00-

***COPD rate(QI #23)***

**COPD**

ICD-9-CM diagnosis codes:
4660 AC BRONCHITIS*
490 BRONCHITIS NOS*
4910 SIMPLE CHR BRONCHITIS
4911 MUCOPURUL CHR BRONCHITIS
49120 OBS CHR BRNC W/O ACT EXA
49121 OBS CHR BRNC W ACT EXA
4918 CHRONIC BRONCHITIS NEC
4919 CHRONIC BRONCHITIS NOS
4920 EMPHYSEMATOUS BLEB
4928 EMPHYSEMA NEC
494 BRONCHIECTASIS -OCT00
4940 BRONCHIECTAS W/O AC EXAC OCT00-
4941 BRONCHIECTASIS W AC EXAC OCT00-
496 CHR AIRWAY OBSTRUCT NEC

***** Qualifies only if accompanied by secondary diagnosis of 491.xx, 492.x, 494, or 496 (i.e., any other code on this list).

***Congestive heart failure rate(QI #24)***

**CHF**

ICD-9-CM diagnosis codes:
39891 RHEUMATIC HEART FAILURE
40201 MAL HYPERT HRT DIS W CHF
40211 BENIGN HYP HRT DIS W CHF
40291 HYPERTEN HEART DIS W CHF
40401 MAL HYPER HRT/REN W CHF
40403 MAL HYP HRT/REN W CHF&RF
40411 BEN HYPER HRT/REN W CHF
40413 BEN HYP HRT/REN W CHF&RF
40491 HYPER HRT/REN NOS W CHF
40493 HYP HT/REN NOS W CHF&RF
4280 CONGESTIVE HEART FAILURE
4281 LEFT HEART FAILURE
4289 HEART FAILURE NOS

*Exclude:* **Cardiac procedures:**

ICD-9-CM procedure codes:
3601 PTCA-1 VESSEL W/O AGENT
3602 PTCA-1 VESSEL WITH AGNT
3605 PTCA-MULTIPLE VESSEL
3606 INSERT CORONARY ART STENT OCT95-
3610 AORTOCORONARY BYPASS NOS
3611 AORTOCOR BYPAS-1 COR ART
3612 AORTOCOR BYPAS-2 COR ART
3613 AORTOCOR BYPAS-3 COR ART
3614 AORTCOR BYPAS-4+ COR ART
3615 1 INT MAM-COR ART BYPASS
3616 2 INT MAM-COR ART BYPASS
3617 ABD-CORON ART BYPASS OCT96-
3619 HRT REVAS BYPS ANAS NEC
375 HEART TRANSPLANTATION
3770 INT INSERT PACEMAK LEAD
3771 INT INSERT LEAD IN VENT
3772 INT INSER LEAD ATRI-VENT
3773 INT INSER LEAD IN ATRIUM
3774 INT OR REPL LEAD EPICAR
3775 REVISION OF LEAD
3776 REPL TV ATRI-VENT LEAD
3777 REMOVAL OF LEAD W/O REPL
3778 INSER TEMP PACEMAKER SYS
3779 REVIS OR RELOCATE POCKET

***Diabetes – short-term complications rate(QI #25)***

**Diabetes with short-term complications**

ICD-9-CM Diagnosis Codes:
25010 DM KETO T2, DM CONT
25011 DM KETO T1, DM CONT
25012 DM KETO T2, DM UNCONT
25013 DM KETO T1, DM UNCONT
25020 DM W/ HYPROSM T2, DM CONT
25021 DM W/ HYPROSM T1, DM CONT
25022 DM W/ HYPROSM T2, DM UNCNT
25023 DM W/ HYPROSM T1, DM UNCNT
25030 DM COMA NEC TYP II, DM CNT
25031 DM COMA NEC T1, DM CONT
25032 DM COMA NEC T2, DM UNCONT
25033 DM COMA NEC T1, DM UNCONT

***Diabetes – uncontrolled (without short-term or long-term complications) admission rate (QI #26)***

**Uncontrolled diabetes**

ICD-9-CM Diagnosis Codes:
25002 DM, T2, UNCONT
25003 DM, T1, UNCONT

***Diabetes – long-term complications rate(QI #27)***

**Diabetes with long term-complications**

ICD-9-CM Diagnosis Codes:
25040 DM RENAL COMP T2 CONT
25041 DM RENAL COMP T1 CONT
25042 DM RENAL COMP T2 UNCNT
25043 DM RENAL COMP T1 UNCNT
25050 DM EYE COMP T2 CONT
25051 DM EYE COMP T1 CONT
25052 DM EYE COMP T2 UNCNT
25053 DM EYE COMP T1 UNCNT
25060 DM NEURO COMP T2 CONT
25061 DM NEURO COMP T1 CONT
25062 DM NEURO COMP T2 UNCNT
25063 DM NEURO COMP T1 UNCNT
25070 DM CIRCU DIS T2 CONT
25071 DM CIRCU DIS T1 CONT
25072 DM CIRCU DIS T2 UNCNT
25073 DM CIRCU DIS T1 UNCNT
25080 DM W COMP NEC T2 CONT
25081 DM W COMP NEC T1 CONT
25082 DM W COMP NEC T2 UNCNT
25083 DM W COMP NEC T1 UNCNT
25090 DM W COMPL NOS T2 CONT
25091 DM W COMPL NOS T1 CONT
25092 DM W COMPL NOS T2 UNCNT
25093 DM W COMPL NOS T1 UNCNT

***Hypertension rate(QI #28)***

**Hypertension**

ICD-9-CM diagnosis codes:
4010 MALIGNANT HYPERTENSION
4019 HYPERTENSION NOS
40200 MAL HYPERTEN HRT DIS W/OUT CHF
40210 BEN HYPERTEN HRT DIS W/OUT CHF

40290 HYPERTENSIVE HRT DIS W/OUT CHF
40300 MAL HYPERT HRT DIS W/OUT RF
40310 BENIGN HYP HRT DIS W/OUT RF
40390 HYPERTEN HEART DIS W/OUT RF
40400 MAL HYPER HRT/REN W/OUT CHF/RF
40410 BEN HYPER HRT/REN W/OUT CHF/RF
40490 HYPER HRT/REN NOS W/OUT CHF/RF

*Exclude:* **Cardiac procedures:**

ICD-9-CM procedure codes:
3601 PTCA-1 VESSEL W/O AGENT
3602 PTCA-1 VESSEL WITH AGNT
3605 PTCA-MULTIPLE VESSEL-
3606 INSERT CORONARY ART STENT OCT95-
3610 AORTOCORONARY BYPASS NOS
3611 AORTOCOR BYPAS-1 COR ART
3612 AORTOCOR BYPAS-2 COR ART
3613 AORTOCOR BYPAS-3 COR ART
3614 AORTCOR BYPAS-4+ COR ART
3615 1 INT MAM-COR ART BYPASS
3616 2 INT MAM-COR ART BYPASS
3617 ABD-CORON ART BYPASS OCT96-
3619 HRT REVAS BYPS ANAS NEC
375 HEART TRANSPLANTATION
3770 INT INSERT PACEMAK LEAD
3771 INT INSERT LEAD IN VENT
3772 INT INSER LEAD ATRI-VENT
3773 INT INSER LEAD IN ATRIUM
3774 INT OR REPL LEAD EPICAR
3775 REVISION OF LEAD
3776 REPL TV ATRI-VENT LEAD
3777 REMOVAL OF LEAD W/O REPL
3778 INSER TEMP PACEMAKER SYS
3779 REVIS OR RELOCATE POCKET

***Lower extremity amputation rate(QI #29)***

**Lower extremity amputation**

ICD-9-CM procedure codes:
8410 LOWER LIMB AMPUTAT NOS
8411 TOE AMPUTATION
8412 AMPUTATION THROUGH FOOT
8413 DISARTICULATION OF ANKLE
8414 AMPUTAT THROUGH MALLEOLI
8415 BELOW KNEE AMPUTAT NEC
8416 DISARTICULATION OF KNEE
8417 ABOVE KNEE AMPUTATION
8418 DISARTICULATION OF HIP
8419 HINDQUARTER AMPUTATION

*Include only:* **Diabetes**

ICD-9-CM diagnosis codes:
25000 DMII WO CMP NT ST UNCNTR
25001 DMI WO CMP NT ST UNCNTRL
25002 DMII WO CMP UNCNTRLD
25003 DMI WO CMP UNCNTRLD
25010 DMII KETO NT ST UNCNTRLD
25011 DMI KETO NT ST UNCNTRLD
25012 DMII KETOACD UNCONTROLD
25013 DMI KETOACD UNCONTROLD
25020 DMII HPRSM NT ST UNCNTRL
25021 DMI HPRSM NT ST UNCNTRLD
25022 DMII HPROSMLR UNCONTROLD
25023 DMI HPROSMLR UNCONTROLD
25030 DMII O CM NT ST UNCNTRLD
25031 DMI O CM NT ST UNCNTRLD
25032 DMII OTH COMA UNCONTROLD
25033 DMI OTH COMA UNCONTROLD
25040 DMII RENL NT ST UNCNTRLD
25041 DMI RENL NT ST UNCNTRLD
25042 DMII RENAL UNCNTRLD
25043 DMI RENAL UNCNTRLD
25050 DMII OPHTH NT ST UNCNTRL
25051 DMI OPHTH NT ST UNCNTRLD
25052 DMII OPHTH UNCNTRLD
25053 DMI OPHTH UNCNTRLD
25060 DMII NEURO NT ST UNCNTRL
25061 DMI NEURO NT ST UNCNTRLD
25062 DMII NEURO UNCNTRLD
25063 DMI NEURO UNCNTRLD
25070 DMII CIRC NT ST UNCNTRLD
25071 DMI CIRC NT ST UNCNTRLD
25072 DMII CIRC UNCNTRLD
25073 DMI CIRC UNCNTRLD
25080 DMII OTH NT ST UNCNTRLD
25081 DMI OTH NT ST UNCNTRLD
25082 DMII OTH UNCNTRLD
25083 DMI OTH UNCNTRLD
25090 DMII UNSPF NT ST UNCNTRL
25091 DMI UNSPF NT ST UNCNTRLD
25092 DMII UNSPF UNCNTRLD
25093 DMI UNSPF UNCNTRLD

*Exclude:* **Trauma**

ICD-9-CM diagnosis codes:
8950 AMPUTATION TOE
8951 AMPUTATION TOE-COMPLICAT
8960 AMPUTATION FOOT, UNILAT
8961 AMPUT FOOT, UNILAT-COMPL
8962 AMPUTATION FOOT, BILAT
8963 AMPUTAT FOOT, BILAT-COMP
8970 AMPUT BELOW KNEE, UNILAT
8971 AMPUTAT BK, UNILAT-COMPL
8972 AMPUT ABOVE KNEE, UNILAT
8973 AMPUT ABV KN, UNIL-COMPL
8974 AMPUTAT LEG, UNILAT NOS
8975 AMPUT LEG, UNIL NOS-COMP
8976 AMPUTATION LEG, BILAT
8977 AMPUTAT LEG, BILAT-COMPL

***Low birthweight rate(QI #30)***

**Low birthweight (Outcome of interest)**

ICD-9-CM diagnosis codes:
76400 LIGHT-FOR-DATES WTNOS
76401 LIGHT-FOR-DATES <500G
76402 LT-FOR-DATES 500-749G
76403 LT-FOR-DATES 750-999G
76404 LT-FOR-DATES 1000-1249G
76405 LT-FOR-DATES 1250-1499G
76406 LT-FOR-DATES 1500-1749G
76407 LT-FOR-DATES 1750-1999G
76408 LT-FOR-DATES 2000-2499G
76410 LT-FOR-DATE W/MAL WTNOS
76411 LT-FOR-DATE W/MAL <500G
76412 LT-DATE W/MAL 500-749G
76413 LT-DATE W/MAL 750-999G
76414 LT-DATE W/MAL 1000-1249G
76415 LT-DATE W/MAL 1250-1499G
76416 LT-DATE W/MAL 1500-1749G
76417 LT-DATE W/MAL 1750-1999G
76418 LT-DATE W/MAL 2000-2499G
76420 FETAL MALNUTRITION WTNOS
76421 FETAL MALNUTRITION <500G
76422 FETAL MALNUTR 500-749G
76423 FETAL MAL 750-999G
76424 FETAL MAL 1000-1249G
76425 FETAL MAL 1250-1499G
76426 FETAL MAL 1500-1749G
76427 FETAL MALNUTR 1750-1999G
76428 FETAL MALNUTR 2000-2499G
76490 FET GROWTH RETARD WTNOS
76491 FET GROWTH RETARD <500G
76492 FET GROWTH RET 500-749G

76493 FET GROWTH RET 750-999G
76494 FET GRWTH RET 1000-1249G
76495 FET GRWTH RET 1250-1499G
76496 FET GRWTH RET 1500-1749G
76497 FET GRWTH RET 1750-1999G
76498 FET GRWTH RET 2000-2499G
76500 EXTREME IMMATUR WTNOS
76501 EXTREME IMMATUR <500G
76502 EXTREME IMMATUR 500-749G
76503 EXTREME IMMATUR 750-999G
76504 EXTREME IMMAT 1000-1249G
76505 EXTREME IMMAT 1250-1499G
76506 EXTREME IMMAT 1500-1749G
76507 EXTREME IMMAT 1750-1999G
76508 EXTREME IMMAT 2000-2499G
76510 PRETERM INFANT NEC WTNOS
76511 PRETERM NEC <5 00G
76512 PRETERM NEC 500-749G
76513 PRETERM NEC 750-999G
76514 PRETERM NEC 1000-1249G
76515 PRETERM NEC 1250-1499G
76516 PRETERM NEC 1500-1749G
76517 PRETERM NEC 1750-1999G
76518 PRETERM NEC 2000-2499G

**All births (pouulation at risk)**

**DRG Codes:**
DRGs:
370 CESAREAN SECTION W CC
371 CESAREAN SECTION W/O CC
372 VAGINAL DELIVERY W COMPL
373 VAG DELIVERY W/O COMPL
374 VAG DELIV W STERIL OR DC
375 VAG DELIV W OTH OR PROC

***Pediatric asthma rate(QI #31)***

**Pediatric asthma**

ICD-9-CM diagnosis codes:
49300 EXT ASTHMA W/O STAT ASTH
49301 EXT ASTHMA W STATUS ASTH
49302 EXT ASTHMA W STATUS ACEX OCT00-
49310 INT ASTHMA W/O STAT ASTH
49311 INT ASTHMA W STATUS ASTH
49312 INT ASTHMA W STATUS ACEX OCT00-
49320 CH OB ASTH W/O STAT ASTH
49321 CH OB ASTHMA W STAT ASTH
49322 CH OB ASTHMA W STAT ACEX OCT00-
49390 ASTHMA W/O STATUS ASTHM
49391 ASTHMA W STATUS ASTHMAT
49392 ASTHMA W STATUS AC EXAC OCT00-

***Pediatric Gastroenteritis rate(QI #32)***

**Pediatric gastroenteritis**

ICD-9-CM diagnosis codes:
00861 ROTAVIRUS ENTERITIS
00862 ADENOVIRUS ENTERITIS
00863 NORWALK ENTERITIS
00864 OTHER SRV ENTERITIS
00865 CALICIVIRUS ENTERITIS
00866 ASTROVIRUS ENTERITIS
00867 ENTEROVIRUS ENTERITIS NEC
00869 ENTERITIS NOS
0088 VIRAL ENTERITIS NOS
0090 INFECTIOUS ENTERITIS NOS
0091 ENTERITIS OF INFECT ORIG
0092 INFECTIOUS DIARRHEA
0093 DIARRHEA, PRESUM INFECTIOUS
5589 NONINF GASTROENTERIT NEC

## Mortality

***AMI mortality (QI #33)***

**AMI**

ICD-9-CM diagnosis codes:
41001 AMI ANTEROLATERAL, INIT
41011 AMI ANTERIOR WALL, INIT
41021 AMI INFEROLATERAL, INIT
41031 AMI INFEROPOST, INITIAL
41041 AMI INFERIOR WALL, INIT
41051 AMI LATERAL NEC, INITIAL
41061 TRUE POST INFARCT, INIT
41071 SUBENDO INFARCT, INITIAL
41081 AMI NEC, INITIAL
41091 AMI NOS, INITIAL

***Congestive heart failure mortality (QI #34)***

**CHF**

ICD-9-CM diagnosis codes:
39891 RHEUMATIC HEART FAILURE
40201 MAL HYPERT HRT DIS W CHF
40211 BENIGN HYP HRT DIS W CHF
40291 HYPERTEN HEART DIS W CHF
40401 MAL HYPER HRT/REN W CHF
40403 MAL HYP HRT/REN W CHF&RF
40411 BEN HYPER HRT/REN W CHF
40413 BEN HYP HRT/REN W CHF&RF
40491 HYPER HRT/REN NOS W CHF
40493 HYP HT/REN NOS W CHF&RF
4280 CONGESTIVE HEART FAILURE
4281 LEFT HEART FAILURE
4289 HEART FAILURE NOS

*Exclude:* **Cardiac procedures:**

ICD-9-CM procedure codes:
3601 PTCA-1 VESSEL W/O AGENT
3602 PTCA-1 VESSEL WITH AGNT
3607 PTCA-MULTIPLE VESSEL
3608 INSERT CORONARY ART STENT OCT95-
3610 AORTOCORONARY BYPASS NOS
3611 AORTOCOR BYPAS-1 COR ART
3612 AORTOCOR BYPAS-2 COR ART
3613 AORTOCOR BYPAS-3 COR ART
3614 AORTCOR BYPAS-4+ COR ART
3615 1 INT MAM-COR ART BYPASS
3616 2 INT MAM-COR ART BYPASS
3617 ABD-CORON ART BYPASS OCT96-
3619 HRT REVAS BYPS ANAS NEC
375 HEART TRANSPLANTATION
3770 INT INSERT PACEMAK LEAD
3771 INT INSERT LEAD IN VENT
3772 INT INSER LEAD ATRI-VENT
3773 INT INSER LEAD IN ATRIUM
3774 INT OR REPL LEAD EPICAR
3775 REVISION OF LEAD
3776 REPL TV ATRI-VENT LEAD
3777 REMOVAL OF LEAD W/O REPL
3778 INSER TEMP PACEMAKER SYS
3779 REVIS OR RELOCATE POCKET

***GI Hemorrhage mortality(QI #35)***

**GI Hemorrhage**

ICD-9-CM diagnosis codes:
4560 ESOPH VARICES W HEM
5307 GASTROESOPH LACER W HEM
53082 ESOPHAGEAL HEM
53100 AC STOMACH ULCER W HEM
53101 AC STOMAC ULC W HEM-OBST
53120 AC STOMAC ULC W HEM/PERF

53121 AC STOM ULC HEM/PERF-OBS
53140 CHR STOMACH ULC W HEM
53141 CHR STOM ULC W HEM-OBSTR
53160 CHR STOMACH ULC HEM/PERF
53161 CHR STOM ULC HEM/PERF-OB
53200 AC DUODENAL ULCER W HEM
53201 AC DUODEN ULC W HEM-OBST
53220 AC DUODEN ULC W HEM/PERF
53221 AC DUOD ULC HEM/PERF-OBS
53240 CHR DUODEN ULCER W HEM
53241 CHR DUODEN ULC HEM-OBSTR
53260 CHR DUODEN ULC HEM/PERF
53261 CHR DUOD ULC HEM/PERF-OB
53300 AC PEPTIC ULCER W HEMORR
53301 AC PEPTIC ULC W HEM-OBST
53320 AC PEPTIC ULC W HEM/PERF
53321 AC PEPT ULC HEM/PERF-OBS
53340 CHR PEPTIC ULCER W HEM
53341 CHR PEPTIC ULC W HEM-OBS
53360 CHR PEPT ULC W HEM/PERF
53361 CHR PEPT ULC HEM/PERF-OB
53400 AC MARGINAL ULCER W HEM
53401 AC MARGIN ULC W HEM-OBST
53420 AC MARGIN ULC W HEM/PERF
53421 AC MARG ULC HEM/PERF-OBS
53440 CHR MARGINAL ULCER W HEM
53441 CHR MARGIN ULC W HEM-OBS
53460 CHR MARGIN ULC HEM/PERF
53461 CHR MARG ULC HEM/PERF-OB
53501 ACUTE GASTRITIS W HMRHG
53511 ATRPH GASTRITIS W HMRHG
53521 GSTR MCSL HYPRT W HMRG
53531 ALCHL GSTRITIS W HMRHG
53541 OTH SPF GASTRT W HMRHG
53551 GSTRDDNTS NOS W HMRHG
53561 DUODENITIS W HMRHG
53783 ANGIODYS STOM DUOD W HEM
56202 DIVERTICULOS SM INT W HEM
56203 DIVERTICULIT SM INT W HEM
56212 DIVERTICULOS COLON W HEM
56213 DIVERTICULIT COLON W HEM
5693 HEMORRHAGE RECT/ANUS
56985 ANGIODYS INTEST W HEM
5780 HEMATEMESIS
5781 BLOOD IN STOOL
5789 GASTROINTEST HEMORR NOS

***Hip fracture mortality(QI #36)***

**Hip fracture**

ICD-9-CM diagnosis codes:
82000 FX FEMUR INTRCAPS NOS-CL
82001 FX UP FEMUR EPIPHY-CLOS
82002 FX FEMUR, MIDCERVIC-CLOS
82003 FX BASE FEMORAL NCK-CLOS
82009 FX FEMUR INT RCAPS NEC-CL
82010 FX FEMUR INTRCAP NOS-OPN
82011 FX UP FEMUR EPIPHY-OPEN
82012 FX FEMUR, MIDCERVIC-OPEN
82013 FX BASE FEMORAL NCK-OPEN
82019 FX FEMUR INTRCAP NEC-OPN
82020 TROCHANTERIC FX NOS-CLOS
82021 INTERTROCHANTERIC FX-CL
82022 SUBTROCHANTERIC FX-CLOSE
82030 TROCHANTERIC FX NOS-OPEN
82031 INTERTROCHANTERIC FX-OPN
82032 SUBTROCHANTERIC FX-OPEN
8208 FX NECK OF FEMUR NOS-CL
8209 FX NECK OF FEMUR NOS-OPN

***Pneumonia mortality(QI #37)***

**Pneumonia**

ICD-9-CM diagnosis codes:
00322 SALMONELLA PNEUMONIA
0212 PULMONARY TULAREMIA
0391 PULMONARY ACTINOMYCOSIS
0521 VARICELLA PNEUMONITIS
0551 POSTMEASLES PNEUMONIA
0730 ORNITHOSIS PNEUMONIA
1124 CANDIDIASIS OF LUNG
1140 PRIMARY COCCIDIOIDOMYCOS
1144 CHRONIC PULMON COCCIDIOIDOMYCOSIS
1145 UNSPEC PULMON COCCIDIOIDOMYCOSIS
11505 HISTOPLASM CAPS PNEUMON
11515 HISTOPLASM DUB PNEUMONIA
11595 HISTOPLASMOSIS PNEUMONIA
1304 TOXOPLASMA PNEUMONITIS
1363 PNEUMOCYSTOSIS
4800 ADENOVIRAL PNEUMONIA
4801 RESP SYNCYT VIRAL PNEUM
4802 PARINFLUENZA VIRAL PNEUM
4808 VIRAL PNEUMONIA NEC
4809 VIRAL PNEUMONIA NOS
481 PNEUMOCOCCAL PNEUMONIA
4820 K. PNEUMONIAE PNEUMONIA
4821 PSEUDOMONAL PNEUMONIA
4822 H.INFLUENZAE PNEUMONIA
48230 STREP PNEUMONIA UNSPEC
48231 GRP A STREP PNEUMONIA
48232 GRP B STREP PNEUMONIA
48239 OTH STREP PNEUMONIA
4824 STAPHYLOCOCCAL PNEUMONIA
48240 STAPH PNEUMONIA UNSP OCT98-
48241 STAPH AUREUS PNEUMON OCT98-
48249 STAPH PNEUMON OTH OCT98-
48281 ANAEROBIC PNEUMONIA
48282 E COLI PNEUMONIA
48283 OTH GRAM NEG PNEUMONIA
48284 LEGIONNAIRES DX OCT97-
48289 BACT PNEUMONIA NEC
4829 BACTERIAL PNEUMONIA NOS
4830 MYCOPLASMA PNEUMONIA
4831 CHLAMYDIA PNEUMONIA OCT96-
4838 OTH SPEC ORG PNEUMONIA
4841 PNEUM W CYTOMEG INCL DIS
4843 PNEUMONIA IN WHOOP COUGH
4845 PNEUMONIA IN ANTHRAX
4846 PNEUM IN ASPERGILLOSIS
4847 PNEUM IN OTH SYS MYCOSES
4848 PNEUM IN INFECT DIS NEC
485 BRONCOPNEUMONIA ORG NOS
486 PNEUMONIA, ORGANISM NOS
5070 PNEUMONITIS, ASP FOOD VOMIT
5100 EMPYEMA W FISTULA
5109 EMPYEMA W/O FISTULA
5110 PLEURISY W EFF, BACT NOT TB
5130 ABSCESS OF LUNG

***Stroke mortality(QI #38)***

**Stroke**

ICD-9-CM diagnosis codes:
430 SUBARACHNOID HEMORRHAGE
431 INTRACEREBRAL HEMORRHAGE
4320 NONTRAUM EXTRADURAL HEM
4321 SUBDURAL HEMORRHAGE
4329 INTRACRANIAL HEMORR NOS
43301 BASI ART OCCL W/ INFARCT
43311 CAROTD OCCL W/ INFRCT
43321 VERTB ART OCCL W/ INFRCT
43331 MULT PRECER OCCL W/ INFRCT
43381 PRECER OCCL NEC W/ INFRCT
43391 PRECER OCCL NOS W/ INFRCT
43401 CERE THROMBOSIS W/ INFRCT

43411 CERE EMBOLISM W/ INFRCT
43491 CEREB OCCL NOS W/ INFRCT
436 CVA

***AAA repair mortality(QI #39)***

**AAA repair**

ICD-9-CM procedure codes:
3834 AORTA RESECTION & ANAST
3844 RESECT ABDM AORTA W REPL
3864 EXCISION OF AORTA

*Include only:* **AAA**

ICD-9-CM diagnosis codes:
4413 RUPT ABD AORTIC ANEURYSM
4414 ABDOM AORTIC ANEURYSM

***CABG Mortality(QI #40)***

**CABG**

ICD-9-CM procedure codes:
3610 AORTOCORONARY BYPASS NOS
3611 AORTOCOR BYPAS-1 COR ART
3612 AORTOCOR BYPAS-2 COR ART
3613 AORTOCOR BYPAS-3 COR ART
3614 AORTCOR BYPAS-4+ COR ART
3615 INT MAM-COR ART BYPASS
3616 INT MAM-COR ART BYPASS
3617 ABD-CORON ART BYPASS OCT96-
3619 HRT REVAS BYPS ANAS NEC

***Esophageal resection Mortality (QI #42)***

**Esophageal resection**

ICD-9-CM procedure code:
4240 ESOPHAGECTOMY NOS
4241 PARTIAL ESOPHAGECTOMY
4242 TOTAL ESOPHAGECTOMY

*Include only*: **Esophageal cancer**:

ICD-9-CM diagnosis codes:
1500 MAL NEO CERVICAL ESOPHAG
1501 MAL NEO THORACIC ESOPHAG
1502 MAL NEO ABDOMIN ESOPHAG
1503 MAL NEO UPPER 3RD ESOPH
1504 MAL NEO MIDDLE 3RD ESOPH
1505 MAL NEO LOWER 3RD ESOPH
1508 MAL NEO ESOPHAGUS NEC
1509 MAL NEO ESOPHAGUS NOS

***Hip Replacement Mortality(QI #43)***

**Hip Replacement**

ICD-9-CM procedure Codes:
8151 TOTAL HIP REPLACEMENT
8152 PARTIAL HIP REPLACEMENT
8153 REVISION HIP REPLACEMENT

*Include only*: **Osteoarthrosis**:

ICD-9-CM diagnosis Codes:
71500 GENL OSTEOARTHROSIS NOS
71509 GENL OSTEOARTHROSIS MULT
71510 LOC PRIM OSTEOART_UNSPEC
71515 LOC PRIM OSTEOART_PELVIS
71518 LOC PRIM OSTEOARTHR NEC
71520 LOC 2ND OSTEOARTH_UNSPEC
71525 LOC 2ND OSTEOARTH_PELVIS
71528 LOC 2ND OSTEOARTHROS NEC
71530 LOC OSTEOARTH NOS_UNSPEC
71535 LOC OSTEOARTH NOS_PELVIS
71538 LOC OSTEOAR NOS_SITE NEC
71580 OSTEOARTHROSIS_MULT SITE
71589 OSTEOARTHROSIS_MULT SITE
71590 OSTEOARTHROS NOS_UNSPEC
71595 OSTEOARTHROS NOS_PELVIS
71598 OSTEOARTHRO NOS_OTH SITE
71650 POLYARTHRITIS NOS UNSPEC
71655 POLYARTHRITIS NOS PELVIS
71658 POLYARTHRITIS NOS NEC
71659 POLYARTHRITIS NOS MULT SITE
71660 MONOARTHRITIS NOS UNSPEC
71665 MONOARTHRITIS NOS PELVIS
71668 MONOARTHRITIS NOS NEC
71669 MONOARTHRITIS NOS MULT SITE
71690 ARTHROPATHY NOS UNSPEC
71695 ARTHROPATHY NOS PELVIS
71698 ARTHROPATHY NOS NEC
71699 ARTHROPATHY NOS MULT SITE

***Pancreatic resection mortality(QI #44)***

**Pancreatic resection**

ICD-9-CM procedure codes:
526 TOTAL PANCREATECTOMY
527 RAD PANCREATICODUODENECT

*Include only*: **Pancreatic cancer**

ICD-9-CM diagnosis codes:
1520 MALIGNANT NEOPL DUODENUM
1561 MAL NEO EXTRAHEPAT DUCTS
1562 MAL NEO AMPULLA OF VATER
1570 MAL NEO PANCREAS HEAD
1571 MAL NEO PANCREAS BODY
1572 MAL NEO PANCREAS TAIL
1573 MAL NEO PANCREATIC DUCT
1574 MAL NEO ISLET LANGERHANS
1578 MALIG NEO PANCREAS NEC
1579 MALIG NEO PANCREAS NOS

***Pediatric heart surgery mortality(QI #45)***

**See pediatric heart surgery volume**

# APPENDIX 7

# Empirical Evaluation Results Tables

This appendix summarizes the empirical results for each indicator in table format. Indicators are in order according to their unique identifier.

## Indicator 1: AAA repair volume

Type: Volume

*Definition: Provider level. Number of AAA Procedures*
*Threshold 1: 10/yr, Threshold 2: 32/yr*

| **Procedure volume:** | | **Number of procedures performed per year** | | | | |
|---|---|---|---|---|---|---|
| | # Hosp with >=1 procedures/yr | Mean | Standard deviation | 50th percentile | 90th percentile | 95th percentile |
| 1995 | 716 | 14 | 16 | 8 | 35 | 47 |
| 1996 | 727 | 14 | 16 | 8 | 36 | 46 |
| 1997 | 717 | 14 | 16 | 8 | 34 | 52 |

| **Procedure share:** | Share of procedures done at high volume hospitals | | Share of hospitals at high volume | |
|---|---|---|---|---|
| | Threshold 1 | Threshold 2 | Threshold 1 | Threshold 2 |
| 1995 | 84.19% | 43.11% | 46.09% | 12.57% |
| 1996 | 83.87% | 42.99% | 44.43% | 12.24% |
| 1997 | 84.70% | 41.73% | 46.44% | 11.58% |

| **Persistence of high volume:** | Share of high volume hospitals remaining high volume | |
|---|---|---|
| | Threshold 1 | Threshold 2 |
| 1995-1996 | 86.24% | 81.11% |
| 1996-1997 | 87.03% | 75.86% |

## Indicator 2: Carotid endarterectomy volume

Type: Volume

*Definition: Provider level. Number of Carotid Endarterectomy Procedures*
*Threshold 1: 50/yr, Threshold 2: 101/yr*

| **Procedure volume:** | | **Number of procedures performed per year** | | | | |
|---|---|---|---|---|---|---|
| | # Hosp with >=1 procedures/yr | Mean | Standard deviation | 50th percentile | 90th percentile | 95th percentile |
| 1995 | 893 | 52 | 60 | 32 | 127 | 158 |
| 1996 | 904 | 52 | 60 | 31.5 | 129 | 169 |
| 1997 | 884 | 54 | 64 | 32 | 133 | 175 |

| **Procedure share:** | Share of procedures done at high volume hospitals | | Share of hospitals at high volume | |
|---|---|---|---|---|
| | Threshold 1 | Threshold 2 | Threshold 1 | Threshold 2 |
| 1995 | 76.08% | 50.32% | 35.05% | 16.35% |
| 1996 | 77.77% | 51.03% | 36.61% | 16.70% |
| 1997 | 78.71% | 52.54% | 37.33% | 17.08% |

| **Persistence of high volume:** | Share of high volume hospitals remaining high volume | |
|---|---|---|
| | Threshold 1 | Threshold 2 |
| 1995-1996 | 93.53% | 87.50% |
| 1996-1997 | 89.72% | 87.76% |

## Indicator 3: CABG volume

Type: Volume

*Definition: Provider level. Number of CABG Procedures*
*Threshold 1: 100/yr, Threshold 2: 200/yr*

| **Procedure volume:** | | **Number of procedures performed per year** | | | | |
|---|---|---|---|---|---|---|
| | # Hosp with >=1 procedures/yr | Mean | Standard deviation | 50th percentile | 90th percentile | 95th percentile |
| 1995 | 304 | 375 | 323 | 280 | 804 | 998 |
| 1996 | 307 | 399 | 338 | 293 | 830 | 1095 |
| 1997 | 307 | 401 | 345 | 301 | 892 | 1074 |

| **Procedure share:** | Share of procedures done at high volume hospitals | | Share of hospitals at high volume | |
|---|---|---|---|---|
| | Threshold 1 | Threshold 2 | Threshold 1 | Threshold 2 |
| 1995 | 98.03% | 89.50% | 86.18% | 65.13% |
| 1996 | 98.28% | 90.65% | 87.62% | 67.75% |
| 1997 | 98.30% | 90.71% | 87.95% | 68.08% |

| **Persistence of high volume:** | Share of high volume hospitals remaining high volume | |
|---|---|---|
| | Threshold 1 | Threshold 2 |
| 1995-1996 | 98.85% | 96.94% |
| 1996-1997 | 98.08% | 97.49% |

## Indicator 4: Esophageal resection volume

Type: Volume

*Definition: Provider level. Number of Esophageal Resection Procedures*
*Threshold 1: 6/yr, Threshold 2: 7/yr*

| **Procedure volume:** | | **Number of procedures performed per year** | | | | |
|---|---|---|---|---|---|---|
| | # Hosp with >=1 procedures/yr | Mean | Standard deviation | 50th percentile | 90th percentile | 95th percentile |
| 1995 | 265 | 2 | 3 | 1 | 4 | 7 |
| 1996 | 265 | 2 | 3 | 1 | 4 | 6 |
| 1997 | 233 | 2 | 3 | 1 | 5 | 8 |

| **Procedure share:** | Share of procedures done at high volume hospitals | | Share of hospitals at high volume | |
|---|---|---|---|---|
| | Threshold 1 | Threshold 2 | Threshold 1 | Threshold 2 |
| 1995 | 30.43% | 26.26% | 6.79% | 5.28% |
| 1996 | 39.51% | 34.26% | 8.58% | 6.43% |
| 1997 | 30.85% | 24.47% | 5.15% | 3.43% |

| **Persistence of high volume:** | Share of high volume hospitals remaining high volume | |
|---|---|---|
| | Threshold 1 | Threshold 2 |
| 1995-1996 | 50.00% | 57.14% |
| 1996-1997 | 58.33% | 60.00% |

## Indicator 5: Pancreatic resection volume

Type: Volume

*Definition: Provider level. Number of Pancreatic Resection Procedures*
*Threshold 1: 10/yr, Threshold 2: 11/yr*

| **Procedure volume:** | | **Number of procedures performed per year** | | | | |
|---|---|---|---|---|---|---|
| | # Hosp with >=1 procedures/yr | Mean | Standard deviation | 50th percentile | 90th percentile | 95th percentile |
| 1995 | 461 | 3 | 5 | 2 | 5 | 8 |
| 1996 | 429 | 3 | 4 | 2 | 5 | 10 |
| 1997 | 453 | 3 | 5 | 2 | 6 | 8 |

| **Procedure share:** | Share of procedures done at high volume hospitals | | Share of hospitals at high volume | |
|---|---|---|---|---|
| | Threshold 1 | Threshold 2 | Threshold 1 | Threshold 2 |
| 1995 | 30.22% | 27.96% | 4.77% | 4.12% |
| 1996 | 30.25% | 27.03% | 5.13% | 4.19% |
| 1997 | 31.58% | 31.58% | 4.41% | 4.41% |

| **Persistence of high volume:** | Share of high volume hospitals remaining high volume | |
|---|---|---|
| | Threshold 1 | Threshold 2 |
| 1995-1996 | 72.73% | 73.68% |
| 1996-1997 | 71.43% | 82.35% |

## Indicator 6: Pediatric heart surgery volume

Type: Volume

*Definition: Provider level. Number of Pediatric Heart Surgery Procedures*
*Threshold 1: 100/yr, Threshold 2: 100/yr*

| **Procedure volume:** | | **Number of procedures performed per year** | | | | |
|---|---|---|---|---|---|---|
| | # Hosp with >=1 procedures/yr | Mean | Standard deviation | 50th percentile | 90th percentile | 95th percentile |
| 1995 | 135 | 52 | 86 | 2 | 145 | 244 |
| 1996 | 126 | 53 | 90 | 2.5 | 149 | 245 |
| 1997 | 126 | 52 | 91 | 2 | 143 | 263 |

| **Procedure share:** | Share of procedures done at high volume hospitals | | Share of hospitals at high volume | |
|---|---|---|---|---|
| | Threshold 1 | Threshold 2 | Threshold 1 | Threshold 2 |
| 1995 | | 72.88% | | 19.26% |
| 1996 | | 75.52% | | 20.63% |
| 1997 | | 74.33% | | 19.05% |

| **Persistence of high volume:** | Share of high volume hospitals remaining high volume | |
|---|---|---|
| | Threshold 1 | Threshold 2 |
| 1995-1996 | | 84.62% |
| 1996-1997 | | 84.00% |

## Indicator 7: PTCA volume

Type: Volume

*Definition: Provider level. Number of PTCA Procedures*
*Threshold 1: 200/yr, Threshold 2: 400/yr*

| **Procedure volume:** | | **Number of procedures performed per year** | | | | |
|---|---|---|---|---|---|---|
| | # Hosp with >=1 procedures/yr | Mean | Standard deviation | 50th percentile | 90th percentile | 95th percentile |
| 1995 | 354 | 379 | 367 | 292 | 803 | 1066 |
| 1996 | 356 | 418 | 400 | 330 | 869 | 1157 |
| 1997 | 354 | 447 | 428 | 338 | 934 | 1290 |

| **Procedure share:** | Share of procedures done at high volume hospitals | | Share of hospitals at high volume | |
|---|---|---|---|---|
| | Threshold 1 | Threshold 2 | Threshold 1 | Threshold 2 |
| 1995 | 94.41% | 72.54% | 65.82% | 37.29% |
| 1996 | 95.70% | 77.01% | 68.82% | 42.13% |
| 1997 | 95.38% | 79.06% | 69.21% | 44.63% |

| **Persistence of high volume:** | Share of high volume hospitals remaining high volume | |
|---|---|---|
| | Threshold 1 | Threshold 2 |
| 1995-1996 | 97.84% | 98.47% |
| 1996-1997 | 97.88% | 96.53% |

| **Indicator 8: Cesarean section delivery rate** | | | | | | |
|---|---|---|---|---|---|---|
| Type: Provider level utilization | | | | | | |
| *Definition: Provider level. Number of cesarean section delivery per 100 deliveries.* | | | | | | |
| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.214 | 0.087 | 0.045 | 0.012 | 0.880 | 0.922 |
| Demographics | - | - | 0.045 | 0.012 | 0.882 | 0.925 |

| **Minimum bias (age-sex):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.956 | 0.054 | 0.724 | 0.793 | 0.026 |

| **Indicator 9: Incidental appendectomy among elderly rate** | | | | | | |
|---|---|---|---|---|---|---|
| Type: Provider level utilization | | | | | | |
| *Definition: Provider level. Number of incidental appendectomies per 100 patients age 65 years or older undergoing intra-abdominal surgery.* | | | | | | |
| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.027 | 0.035 | 0.019 | 0.015 | 0.554 | 0.670 |
| Demographics | - | - | 0.019 | 0.018 | 0.590 | 0.689 |
| APR-DRG | - | - | 0.019 | 0.014 | 0.554 | 0.673 |

| **Minimum bias (APR-DRG):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.988 | 0.050 | 0.829 | 0.946 | 0.003 |

## Indicator 10: Bi-lateral cardiac catheterization rate

Type: Provider level utilization

*Definition: Provider level. Number of bilateral cardiac catheterizations per 100 heart catheterizations.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.193 | 0.200 | 0.163 | 0.147 | 0.944 | 0.963 |
| Demographics | - | - | 0.162 | 0.140 | 0.942 | 0.961 |
| APR-DRG | - | - | 0.161 | 0.144 | 0.943 | 0.962 |

| **Minimum bias (APR-DRG):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.988 | 0.061 | 0.707 | 0.966 | 0.002 |

## Indicator 11: Successful vaginal birth after cesarean section (VBAC) rate

Type: Provider level utilization

*Definition: Provider level. Number of successful vaginal deliveries per 100 deliveries in women with previous cesarean section.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.336 | 0.148 | 0.115 | 0.055 | 0.825 | 0.889 |
| Demographics | - | - | 0.117 | 0.057 | 0.831 | 0.895 |

| **Minimum bias (age-sex):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.995 | 0.031 | 0.931 | 0.943 | 0.000 |

## Indicator 12: Laparoscopic cholecystectomy rate

Type: Provider level utilization

*Definition: Provider level. Number of laparoscopic cholecystectomies per 100 cholecystectomies (open and laparoscopic) performed.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.662 | 0.192 | 0.137 | 0.083 | 0.836 | 0.893 |
| Demographics | - | - | 0.133 | 0.079 | 0.831 | 0.891 |

| **Minimum bias (age-sex):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.966 | 0.039 | 0.938 | 0.788 | 0.025 |

## Indicator 13: Coronary artery bypass graft (CABG) rate

Type: Area level utilization

*Definition: Area level. Number of CABGs per 100,000 population.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.001804 | 0.005716 | 0.002508 | 0.002152 | 0.973 | 0.974 |
| Demographics | - | - | 0.002486 | 0.002117 | 0.973 | 0.973 |

| **Minimum bias (age-sex):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.6548 | 0.126 | 0.364 | 0.955 | 0.350 |

## Indicator 14: Hysterectomy rate

Type: Area level utilization

*Definition: Area level. Number of hysterectomies per 100,000 population.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.004194 | 0.003233 | 0.001840 | 0.000940 | 0.933 | 0.935 |
| Demographics | - | - | 0.001881 | 0.000983 | 0.936 | 0.937 |

| **Minimum bias (age-sex):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.9964 | 0.030 | 0.818 | 0.909 | 0.000 |

## Indicator 15: Laminectomy and/or spinal fusion rate

Type: Area level utilization

*Definition: Area level. Number of laminectomies per 100,000 population.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.001390 | 0.003475 | 0.001328 | 0.001020 | 0.967 | 0.967 |
| Demographics | - | - | 0.001320 | 0.001009 | 0.967 | 0.967 |

| **Minimum bias (age-sex):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.9333 | 0.063 | 0.318 | 0.955 | 0.074 |

## Indicator 16: Percutaneous transluminal coronary angioplasty (PTCA) rate

Type: Area level utilization

*Definition: Area level. Number of PTCAs per 100,000 population.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.001908 | 0.004556 | 0.002780 | 0.002163 | 0.973 | 0.974 |
| Demographics | - | - | 0.002750 | 0.002120 | 0.973 | 0.973 |

| **Minimum bias (age-sex):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.6712 | 0.104 | 0.364 | 0.955 | 0.355 |

## Indicator 17: Dehydration admission rate

Type: Avoidable hospitalization/ Ambulatory Care Sensitive Condition (ACSC)

*Definition: Area level. Number of admissions for dehydration per 100,000 population.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.001399 | 0.001032 | 0.000484 | 0.000216 | 0.900 | 0.903 |
| Demographics | - | - | 0.000446 | 0.000184 | 0.885 | 0.889 |

| **Minimum bias (age-sex):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.9572 | 0.092 | 0.818 | 0.909 | 0.046 |

## Indicator 18: Bacterial pneumonia admission rate

Type: Avoidable hospitalization/ Ambulatory Care Sensitive Condition (ACSC)

*Definition: Area level. Number of admissions for bacterial pneumonia per 100,000 population.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.003956 | 0.002085 | 0.000971 | 0.000343 | 0.932 | 0.933 |
| Demographics | - | - | 0.000948 | 0.000326 | 0.929 | 0.931 |

| **Minimum bias (age-sex):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.9234 | 0.111 | 0.682 | 0.909 | 0.097 |

## Indicator 19: Urinary infection admission rate

Type: Avoidable hospitalization/ Ambulatory Care Sensitive Condition (ACSC)

*Definition: Area level. Number of admissions for urinary infection per 100,000 population.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.001451 | 0.000895 | 0.000395 | 0.000134 | 0.854 | 0.859 |
| Demographics | - | - | 0.000386 | 0.000128 | 0.849 | 0.854 |

| **Minimum bias (age-sex):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.9143 | 0.115 | 0.682 | 0.909 | 0.088 |

## Indicator 20: Perforated appendix admission rate

Type: Avoidable hospitalization/ Ambulatory Care Sensitive Condition (ACSC)

*Definition: Area level. Number of admissions for perforated appendix per 100 admissions for appendicitis within an area .*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.332740 | 0.144097 | 0.032117 | 0.004604 | 0.308 | 0.436 |
| Demographics | - | - | 0.027529 | 0.003387 | 0.265 | 0.394 |

| **Minimum bias (age-sex):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.9687 | 0.020 | 0.900 | 0.850 | 0.025 |

## Indicator 21: Angina (without procedure) admission rate

Type: Avoidable hospitalization/ Ambulatory Care Sensitive Condition (ACSC)

*Definition: Area level. Number of admissions for angina (without procedure) per 100,000 population.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.001660 | 0.001357 | 0.000639 | 0.000392 | 0.921 | 0.923 |
| Demographics | - | - | 0.000614 | 0.000362 | 0.916 | 0.919 |

| **Minimum bias (age-sex):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.9677 | 0.106 | 0.636 | 0.909 | 0.032 |

## Indicator 22: Adult asthma admission rate.

Type: Avoidable hospitalization/ Ambulatory Care Sensitive Condition (ACSC)

*Definition: Area level. Number of admissions for adult asthma per 100,000 population.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.001079 | 0.000817 | 0.000443 | 0.000158 | 0.840 | 0.846 |
| Demographics | - | - | 0.000435 | 0.000153 | 0.836 | 0.842 |

| **Minimum bias (age-sex):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.9879 | 0.038 | 0.864 | 0.955 | 0.000 |

## Indicator 23: Chronic obstructive pulmonary disease (COPD) admission rate

Type: Avoidable hospitalization/ Ambulatory Care Sensitive Condition (ACSC)

*Definition: Area level. Number of admissions for COPD per 100,000 population.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.003240 | 0.002038 | 0.001153 | 0.000646 | 0.950 | 0.950 |
| Demographics | - | - | 0.000986 | 0.000472 | 0.934 | 0.935 |

| **Minimum bias (age-sex):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.9333 | 0.135 | 0.682 | 0.864 | 0.069 |

## Indicator 24: Congestive heart failure admission rate

Type: Avoidable hospitalization/ Ambulatory Care Sensitive Condition (ACSC)

*Definition: Area level. Number of admissions for CHF per 100,000 population.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.005210 | 0.002865 | 0.001772 | 0.000708 | 0.954 | 0.954 |
| Demographics | - | - | 0.001404 | 0.000444 | 0.930 | 0.932 |

| **Minimum bias (age-sex):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.8577 | 0.196 | 0.545 | 0.818 | 0.166 |

## Indicator 25: Diabetes – short term complications admissions rate

Type: Avoidable hospitalization/ Ambulatory Care Sensitive Condition (ACSC)

*Definition: Area level. Number of admissions for diabetes short term complications (uncontrolled diabetes, DKA, hyperosmolarity, and coma) per 100,000 population.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.000360 | 0.000246 | 0.000092 | 0.000022 | 0.518 | 0.543 |
| Demographics | - | - | 0.000092 | 0.000022 | 0.517 | 0.543 |

| **Minimum bias (age-sex):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.9951 | 0.006 | 1.000 | 1.000 | 0.005 |

## Indicator 26: Uncontrolled diabetes

Type: Avoidable hospitalization/ Ambulatory Care Sensitive Condition (ACSC)

*Definition: Area level. Number of admissions for uncontrolled diabetes per 100,000 population.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.000347 | 0.000281 | 0.000152 | 0.000078 | 0.735 | 0.751 |
| Demographics | - | - | 0.000147 | 0.000074 | 0.726 | 0.742 |

| **Minimum bias (age-sex):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.9764 | 0.079 | 0.773 | 0.909 | 0.023 |

## Indicator 27: Diabetes – long-term complications admission rate

Type: Avoidable hospitalization/ Ambulatory Care Sensitive Condition (ACSC)

*Definition: Area level. Number of admissions for diabetes long-term complications (neurological, circulatory, eye, kidney, NOS), per 100,000 population.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.000808 | 0.000581 | 0.000265 | 0.000077 | 0.741 | 0.752 |
| Demographics | - | - | 0.000279 | 0.000085 | 0.756 | 0.766 |

| **Minimum bias (age-sex):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.9264 | 0.124 | 0.727 | 0.909 | 0.083 |

## Indicator 28: Hypertension admission rate

Type: Avoidable hospitalization/ Ambulatory Care Sensitive Condition (ACSC)

*Definition: Area level. Number of admissions for hypertension per 100,000 population.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.000371 | 0.000322 | 0.000142 | 0.000066 | 0.716 | 0.728 |
| Demographics | - | - | 0.000138 | 0.000060 | 0.699 | 0.712 |

| **Minimum bias (age-sex):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.9631 | 0.091 | 0.727 | 1.000 | 0.032 |

## Indicator 29: Lower extremity amputation in diabetics admission rate

Type: Avoidable hospitalization/ Ambulatory Care Sensitive Condition (ACSC)

*Definition: Area level. Number of admissions for lower extremity amputation and diagnosis code for diabetes per 100,000 population.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.000305 | 0.000427 | 0.000146 | 0.000064 | 0.706 | 0.722 |
| Demographics | - | - | 0.000137 | 0.000057 | 0.685 | 0.702 |

| **Minimum bias (age-sex):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.9192 | 0.131 | 0.591 | 0.909 | 0.074 |

## Indicator 30: Low birth weight rate

Type: Avoidable hospitalization/ Ambulatory Care Sensitive Condition (ACSC)

*Definition: Area level. Number of infants with a birth weight of under 2500 grams per 100 births within area.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.039108 | 0.023135 | 0.011832 | 0.002720 | 0.671 | 0.812 |
| Demographics | - | - | 0.011836 | 0.002722 | 0.671 | 0.812 |

| **Minimum bias:** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | NA | NA | NA | NA | NA |

## Indicator 31: Pediatric asthma admission rate

Type: Avoidable hospitalization/ Ambulatory Care Sensitive Condition (ACSC)

*Definition: Area level. Number of admissions for asthma in patients age 18 years and younger per 100,000 population.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.001541 | 0.001439 | 0.001137 | 0.000514 | 0.853 | 0.858 |
| Demographics | - | - | 0.001129 | 0.000506 | 0.851 | 0.856 |

| **Minimum bias (age-sex):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.9938 | 0.053 | 1.000 | 0.955 | 0.000 |

## Indicator 32: Pediatric gastroenteritis admission rate

Type: Avoidable hospitalization/ Ambulatory Care Sensitive Condition (ACSC)

*Definition: Area level. Number of admissions for gastroenteritis in patients age 18 years and younger per 100,000 population.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.000985 | 0.001012 | 0.000535 | 0.000291 | 0.778 | 0.788 |
| Demographics | - | - | 0.000537 | 0.000293 | 0.778 | 0.788 |

| **Minimum bias (age-sex):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.9953 | 0.055 | 0.909 | 1.000 | 0.000 |

## Indicator 33: Acute myocardial infarction (AMI) mortality rate

Type: In-patient mortality

*Definition: Provider level. Number of deaths per 100 patients with diagnosis of AMI.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.244 | 0.161 | 0.078 | 0.043 | 0.718 | 0.826 |
| Demographics | - | - | 0.061 | 0.023 | 0.617 | 0.756 |
| APR-DRG | - | - | 0.034 | 0.008 | 0.428 | 0.590 |

| **Minimum bias (APR-DRG):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.747 | 0.324 | 0.363 | 0.673 | 0.290 |

## Indicator 34: Congestive heart failure (CHF) mortality rate

Type: In-patient mortality

*Definition: Provider level. Number of deaths per 100 patients with diagnosis of CHF.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.075 | 0.095 | 0.025 | 0.010 | 0.598 | 0.750 |
| Demographics | - | - | 0.021 | 0.007 | 0.505 | 0.696 |
| APR-DRG | - | - | 0.021 | 0.007 | 0.535 | 0.697 |

| **Minimum bias (APR-DRG):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.794 | 0.137 | 0.398 | 0.729 | 0.235 |

## Indicator 35: Gastrointestinal (GI) hemorrhage mortality rate

Type: In-patient mortality

*Definition: Provider level. Number of deaths per 100 patients with diagnosis of GI hemorrhage.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.046 | 0.057 | 0.016 | 0.006 | 0.323 | 0.589 |
| Demographics | - | - | 0.013 | 0.003 | 0.224 | 0.607 |
| APR-DRG | - | - | 0.011 | 0.003 | 0.202 | 0.555 |

| **Minimum bias (APR-DRG):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.803 | 0.105 | 0.504 | 0.678 | 0.229 |

## Indicator 36: Hip fracture mortality rate

Type: In-patient mortality

*Definition: Provider level. Number of deaths per 100 patients with diagnosis of hip fracture.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.144 | 0.160 | 0.079 | 0.062 | 0.544 | 0.670 |
| Demographics | - | - | 0.072 | 0.048 | 0.476 | 0.635 |
| APR-DRG | - | - | 0.078 | 0.060 | 0.543 | 0.653 |

| **Minimum bias (APR-DRG):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.880 | 0.225 | 0.556 | 0.788 | 0.151 |

## Indicator 37: Pneumonia mortality rate

Type: In-patient mortality

*Definition: Provider level. Number of deaths per 100 patients with diagnosis of pneumonia.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.138 | 0.102 | 0.052 | 0.022 | 0.744 | 0.850 |
| Demographics | - | - | 0.045 | 0.015 | 0.675 | 0.822 |
| APR-DRG | - | - | 0.037 | 0.012 | 0.629 | 0.781 |

| **Minimum bias (APR-DRG):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.596 | 0.219 | 0.306 | 0.603 | 0.408 |

## Indicator 38: Acute stroke mortality rate

Type: In-patient mortality

*Definition: Provider level. Number of deaths per 100 patients with diagnosis of stroke (not TIA).*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.213 | 0.137 | 0.070 | 0.029 | 0.626 | 0.756 |
| Demographics | - | - | 0.067 | 0.025 | 0.589 | 0.738 |
| APR-DRG | - | - | 0.053 | 0.017 | 0.519 | 0.707 |

| **Minimum bias (APR-DRG):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.803 | 0.131 | 0.638 | 0.629 | 0.244 |

## Indicator 39: Abdominal aortic aneurysm repair mortality rate

Type: Post-procedural mortality

*Definition: Provider level. Number of deaths per 100 patients having undergone AAA repair.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.215 | 0.268 | 0.085 | 0.060 | 0.423 | 0.540 |
| Demographics | - | - | 0.067 | 0.035 | 0.037 | 0.571 |
| APR-DRG | - | - | NA | NA | NA | NA |

| **Minimum bias (age-sex):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.890 | 0.234 | 0.708 | 0.646 | 0.141 |

## Indicator 40: Coronary artery bypass graft (CABG) mortality rate

Type: Post-procedural mortality

*Definition: Provider level. Number of deaths per 100 patients having undergone CABG.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.051 | 0.062 | 0.017 | 0.007 | 0.611 | 0.726 |
| Demographics | - | - | 0.016 | 0.006 | 0.552 | 0.727 |
| APR-DRG | - | - | 0.014 | 0.005 | 0.545 | 0.698 |

| **Minimum bias (APR-DRG):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.743 | 0.152 | 0.414 | 0.655 | 0.301 |

## Indicator 41: Craniotomy mortality rate

Type: Post-procedural mortality

*Definition: Provider level. Number of deaths per 100 patients having undergone craniotomy.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.162 | 0.185 | 0.084 | 0.071 | 0.589 | 0.719 |
| Demographics | - | - | 0.077 | 0.054 | 0.527 | 0.704 |
| APR-DRG | - | - | 0.037 | 0.015 | 0.289 | 0.490 |

| **Minimum bias (APR-DRG):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.786 | 0.338 | 0.296 | 0.611 | 0.274 |

## Indicator 42: Esophageal resection mortality rate

Type: Post-procedural mortality

*Definition: Provider level. Number of deaths per 100 patients having undergone esophageal resection for cancer.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.202 | 0.366 | 0.043 | 0.026 | 0.120 | 0.353 |
| Demographics | - | - | 0.024 | 0.008 | 0.089 | 0.210 |
| APR-DRG | - | - | NA | NA | NA | NA |

| **Minimum bias (age-sex):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.858 | 0.168 | 0.667 | 1.000 | 0.095 |

## Indicator 43: Hip replacement mortality rate

Type: Post-procedural mortality

*Definition: Provider level. Number of deaths per 100 patients having undergone total or partial hip arthroplasty.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.012 | 0.057 | 0.013 | 0.029 | 0.339 | 0.346 |
| Demographics | - | - | 0.012 | 0.021 | 0.202 | 0.218 |
| APR-DRG | - | - | 0.009 | 0.012 | 0.200 | 0.216 |

| **Minimum bias:** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.642 | 0.489 | 0.278 | 0.708 | 0.361 |

## Indicator 44: Pancreatic resection mortality rate

Type: Post-procedural mortality

*Definition: Provider level. Number of deaths per 100 patients having undergone pancreatic resection for cancer.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.154 | 0.313 | 0.079 | 0.084 | 0.315 | 0.359 |
| Demographics | - | - | 0.042 | 0.032 | 0.165 | 0.347 |
| APR-DRG | - | - | NA | NA | NA | NA |

| **Minimum bias (age-sex):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.540 | 0.419 | 0.714 | 0.286 | 0.381 |

## Indicator 45: Pediatric heart surgery mortality rate

Type: Post-procedural mortality

*Definition: Provider level. Number of deaths per 100 patients age 18 and under having undergone heart surgery.*

| **Precision:** | *Provider Level* | | *Provider Signal* | | | |
|---|---|---|---|---|---|---|
| | Mean | Std Dev | Std Dev (1) | Share (2) | Univariate Ratio (3) | MSX $R^2$ (4) |
| Unadjusted | 0.072 | 0.170 | 0.022 | 0.008 | 0.357 | 0.496 |
| Demographics | - | - | 0.017 | 0.005 | 0.263 | 0.430 |
| APR-DRG | - | - | 0.015 | 0.003 | 0.222 | 0.379 |

| **Minimum bias (APR-DRG):** | Rank Corr. (1) | Absolute Diff (2) | High 10% (3) | Low 10% (4) | Two Decile Diff (5) |
|---|---|---|---|---|---|
| MSX | 0.674 | 0.128 | 0.167 | 0.667 | 0.351 |

# APPENDIX 8

## Literature Tables for Utilization and ACSC Indicators

This appendix summarizes the literature reviews for area utilization and ACSC indicators in table format.

Table 1A. Studies of appropriateness of specific procedures. This table summarizes the studies used to identify area utilization indicators.

Tables 2A-6A. These tables summarize studies of ACSC indicators. Table 2A identifies the studies and designs. Table 3A lists the ACS conditions examined in each study. Tables 4A-6A list pediatric avoidable hospitalizations, infant discretionary hospitalizations, and late hospitalization indicators examined by each study.

Table 3A focuses on conditions for which the risk of hospitalization can be reduced, either through better outpatient management of chronic diseases (e.g., asthma, CHF, diabetes) or through more timely diagnosis and effective treatment of acute conditions (e.g., pneumonia, UTI, cellulitis). This is the best validated of the 4 constructs, and is the basis for all of our selected indicators.

Table 4A includes "conditions for which evidence exists that specific ambulatory care modalities reduce hospitalization rates." This category differs from #1 in that it is more sharply focused on defects in ambulatory care, such as a lack of prior outpatient visits or antibiotic prescriptions. There is some overlap between this list and #1 (e.g., asthma, gastroenteritis, DKA, severe ENT infections, PID), although the definitions often differ slightly. Some of these indicators cannot be implemented without linked outpatient claims.

Table 5A focuses on conditions "for which the decision to admit involves a substantial amount of physician judgment...have a wide range in severity and are often managed at home." This concept seems somewhat less relevant to discussions about quality of care.

Table 6A focuses on conditions "whose advanced stages are presumed to have a greater likelihood of reflecting untimely hospital admissions" (because earlier admission would have prevented progression to the advanced stage). This concept relates to the timeliness of hospitalization and the appropriateness of inpatient care, more than to the timeliness and effectiveness of outpatient care.

**Table 1A. Studies of appropriateness of specific procedures.**

| Procedure | % inappropriate | % uncertain | Source of population | # of patients evaluated |
|---|---|---|---|---|
| Carotid Endarterectomy [1] | 18% (51/281) overall; neurosurgery 14% vs. non-neurosurgery 21%; varied from 0% to 33% among surgeons (P = 0.07) | 49% (138/281) overall, decreased to 45% after adjusting for benefit of CEA for severe symptomatic disease found in NASCET; 40% neurosurgery vs. 55% non-neurosurgery; varied from 33% to 67% across surgeons (P=.26) | All in Edmonton, Alberta, Canada | 291 cases of CEA performed on 265 patients between April 1994 and Sept 1995, from nine surgeons at four teaching hospitals (2 were tertiary-care centers); excluded patients without angiograms (10) |
| Carotid Endartectomy [24] *follow up to above study | 4% (8/184) | 47% (84/184) | All in Edmonton, Alberta, Canada | 184 patients with CEA between 9/1/96 and 8/31/97 were evaluated after results of previous study, CEA guidelines and notification of possible surveillance were distributed to all surgeons performing CEA in Edmonton |
| Carotid Endarterectomy [2] | Definition A (low risk of stroke/death):~55%<br>Definition B: ~5%<br>Definition C:~5% | Definition A (low risk of stroke/death): ~12%<br>Definition B: ~37%<br>Definition C:~14% | Twelve academic medical centers | 1160 randomly selected patients with CEA from 1988-1990 (with the exception of one hospital which included 1987 data), miscoded charts were excluded |
| Carotid Endarterectomy [3] | 32% overall; varied from 29% to 40% among sites | 32% overall; varied from 29% to 34% among sites | 5 sites of varying utilization for the 3 procedures selected from Medicare claims submitted by physicians in Arkansas, Colorado, Iowa, Mass., Montana, Penn., S. Carolina, and N. Calif. | Random sample of Medicare beneficiaries for each procedure (claims submitted in 1981) at each site (high, average, and low use geographic areas) |
| Carotid Endarterectomy [4] | Varied by county from 0% - 67% | No discussion of equivocal indication | 23 adjacent rural and urban, large and small, counties in one large, populous state | Sampled procedures by Medicare billing codes performed on 600 CEA patients in 1981, aged 65 years and older |
| Carotid Endarterectomy [22] | 3.9% characterized as inappropriate; study considered CEA inappropriate if the case was “uncertain” or “proven inappropriate”. | | 1993 Medicare admissions in Georgia w/procedure code for CEA | 1945 CEAs performed on Medicare recipients in GA in 1993 |

| Procedure | % inappropriate | % uncertain | Source of population | # of patients evaluated |
|---|---|---|---|---|
| Cataract Surgery[5] | 2% (15/723) overall; varied from 0% to 6% by institution | 8% (359/723) overall; varied from 0% to 15% by institution | Ten Academic Medical Centers | 1139 randomly selected until approx. 130 patients at each facility w/cataract surgery in 1990 were obtained; patients receiving other ocular surgery performed at the same time as cataract surgery or with specific ICD-9 CM or CPT-4 were excluded |
| Cholecystectomy[6] | 12% overall; varied from 6%-14% (p=.002) among hospitals | 17% overall; varied from 9%-24% (p=.002) among hospitals | Four Israeli hospitals belonging to the General Sick Fund (provides prepaid healthcare to 76% of Israeli population) | 816 patients identified as having undergone cholecystectomy in 1986; 702 records were located and evaluated; complete clinical info was obtained on 657 patients |
| Colonoscopy[7] | 27.8% (110/553) by ASGE criteria; 31.5% (170/553) by US 94 criteria; 25.6% (138/553) by Swiss 94 criteria | No rating for ASGE; 10.9% (59/553) by US 94 criteria; 11.6% (63/553) by Swiss 94 criteria | Two university-based multi-specialty outpatient clinics in Lausanne and Basel, Switzerland | 553 consecutive patients referred by the outpatient clinics for colonoscopy, aged >15 from January 1995 to September 1995 (Lausanne) and January 1995 to July 1995 (Basel) |
| Coronary Angiography[26] | 7% (1/14) of blacks; 10% (4/41) whites who underwent angiography | 50% (7/14) of blacks; 46.3% (19/41) whites who underwent angiography | Department of Veterans Affairs | 200 (100 white and 100 black) VA inpatients discharged between 1/193 and 12/1/93 with primary dx of cardiovascular disease or chest pain |
| Coronary Angiography[8] | 6% overall; no difference across subgroups | 16% overall; no difference across subgroups | Harvard Community Health Plan (HCHP), Brookline Mass; mixed model HMO | 292 HCHP enrollees with coronary angiography in 1992; stratified into four subgroups |
| Coronary Angiography[9] | 21% overall | 30% overall | Trent region; coronary angiography is done in 3 referral centers and CABG in 2 centers. | random sample of 320 patients with coronary angiography between 2/1/87 and 5/30/88. Exclusions: incomplete records, congenital heart disease, transplant, primary valve disease. |

| Procedure | % inappropriate | % uncertain | Source of population | # of patients evaluated |
|---|---|---|---|---|
| Coronary Angiography[3] | 17% overall; varied from 15%-18% among sites | 9% overall; varied from 4%-10% among sites | 5 sites of varying utilization for the 3 procedures selected from Medicare claims submitted by physicians in Arkansas, Colorado, Iowa, Mass., Montana, Penn., S. Carolina, and N. Calif. | Random sample of Medicare beneficiaries for each procedure (claims submitted in 1981) at each site (high, average, and low use geographic areas) |
| Coronary Angiography[4] | Varied by county from 8%-75% | no discussion of equivocal indication | 23 adjacent rural and urban, large and small, counties in one large, populous state | Sampled procedures by Medicare billing codes performed on 600 CA patients in 1981, aged 65 years and older |
| Coronary Angiography[10] | **Canadian Criteria**<br>Canadian sample: 9.0% (95% CI, 6.6%-11.4%)<br>New York sample: 10.2% (95% CI, 8.5%-11.8%)<br>**US Criteria**<br>Canadian sample: 5.1% (95% CI, 3.2%-6.9%)<br>New York sample: 4.2% (95% CI, 3.4%-6.9%) | **Canadian Criteria**<br>Canadian sample: 33.2% (95% CI, 29.2%-37.2%)<br>New York sample: 39.1 (95% CI, 35.1%-43.1%)<br>**US Criteria**<br>Canadian sample: 18.2% (95% CI, 14.9%-21.5%)<br>New York sample: 20.1% (95% CI, 18.4%-21.8%) | All hospitals performing CA and CABG in Ontario and British Columbia; 15 randomly selected hospitals that provide CA in New York State; 15 randomly selected hospitals that provide CABG in New York State | 553 randomly selected patients in Canada, 1333 randomly selected patients in New York. New York patients had procedures performed in 1990; Canadian patients had procedures performed between 4/89 and 3/90. Cases performed primarily for valve surgery were excluded |
| Coronary Angiography[11] | 4% overall; varied from 0% - 9% among hospitals (NS) | 20% overall; varied from 13%-31% among hospitals (NS) | 15 randomly selected, non federal hospitals in New York State providing coronary angiography | Random sample of 1335 patients undergoing angiography in New York State in 1990, distributed across the 15 hospitals |
| CABG[12] | 6% (5/85), compared to 1/85 identified by the original panel of NY cardiologists | 12% (10/85), compared to 1/85 identified by the original panel of NY cardiologists | A follow-up to the above study was done using a sub-sample of the patients. A panel of Duke University cardiologists reviewed 308 records for appropriateness | |
| CABG[13] | 1.6% (95% CI, 0.6% - 2.5%) overall; increased to 1.9% when revised by Consortium surgeons. Varied from 0% to 5% across hospitals (P=0.02) (NS) | 7% (95% CI, 5%-8%) overall; did not vary significantly across hospitals | All 12 Academic Medical Center Consortium hospitals | 1156 patients w/CABG surgery in 1990 w/o previous CABG or concurrent valve replacement surgery, randomly selected consecutively until 100 records were obtained from each facility |

| Procedure | % inappropriate | % uncertain | Source of population | # of patients evaluated |
|---|---|---|---|---|
| CABG[14] | 2.4% (95% CI, 2% - 3%) overall; varied from 0% to 5% among hospitals (NS) | 7% (95% CI, 5%-9%) overall; varied from 3% to 15% among hospitals (NS) | 15 randomly selected, non federal hospitals in New York State providing CABG surgery | Random sample of 1338 patients undergoing isolated CABG in NY in 1990; those undergoing another major procedure in conjunction with CABG (55) were excluded; records missing critical data (13) were also excluded |
| CABG[9] | 16% overall | 26% overall | Trent region; coronary angiography is done in 3 referral centers and CABG in 2 centers. | 319 randomly selected patients with CABG between 7/1/87 and 6/31/88. Exclusions: incomplete records, congenital heart disease, transplant, primary valve disease |
| CABG[10] | **Canadian Criteria**<br>Canadian sample: 3.6% (95% CI, 2.0%-5.1%)<br>New York sample: 5.5% (95% CI, 4.0%-7.1%)<br>**US Criteria**<br>Canadian sample: 2.5% (95% CI, 1.2%-3.8%)<br>New York sample: 2.4% (95% CI, 1.6%-3.1%) | **Canadian Criteria**<br>Canadian sample: 11.3% (95% CI, 8.7%-14.0%)<br>New York sample: 9.9% (95% CI, 8.4%-11.4%)<br>**US Criteria**<br>Canadian sample: 9.0% (95% CI, 6.6%-11.4%)<br>New York sample: 7.0% (95% CI, 5.1%-9.0%) | All hospitals performing CA and CABG in Ontario and British Columbia; 15 randomly selected hospitals that provide CA in New York; 15 randomly selected hospitals that provide CABG in New York | 556 randomly selected CABG patients in Canada, 1336 randomly selected CABG patients in New York. New York patients had procedures performed in 1990; Canadian patients had procedures performed between 4/89 and 3/90. Cases performed primarily for valve surgery were excluded |
| CABG (referral after Coronary Angiography)[16] | 9.7% overall | 12.3% overall | Seven of eight public Swedish heart centers. (perform 92% of all bypass surgeries in Sweden) | Consecutive series of 2767 patients with coronary angiography between 5/94 and 1/95 who were considered for coronary revascularization |
| CABG[15] | **RAND criteria:** 42%<br>**ACC/AHA criteria:** 17%<br>**RAS criteria:** 46% | **RAND criteria:** 17%<br>**ACC/AHA criteria:** no rating<br>**RAS criteria:** no rating | An academic medical center cardiac catheterization laboratory and a VA cardiac catheterization lab in Maryland | 153 catheterization patients referred to a either Univ. of Maryland Cardiac Catheterization Lab and/or Baltimore VA Medical Center Cardiac Catheterization Lab with a variety of cardiac diagnoses and treatments between 3/93 and 10/94 |

| Procedure | % inappropriate | % uncertain | Source of population | # of patients evaluated |
|---|---|---|---|---|
| PTCA[15] | **RAND criteria:** 22%<br>**ACC/AHA criteria:** 49%<br>**RAS criteria:** 35% | **RAND criteria:** 29%<br>**ACC/AHA criteria:** no rating<br>**RAS criteria:** no rating | An academic medical center cardiac catheterization laboratory and a VA cardiac catheterization lab in Maryland | 153 catheterization patients referred to a either Univ. of Maryland Cardiac Catheterization Lab and/or Baltimore VA Medical Center Cardiac Catheterization Lab with a variety of cardiac diagnoses and treatments between 3/93 and 10/94 |
| PTCA (referral after Coronary Angiography)[16] | 38.3% overall | 30.0% overall | Seven of eight public Swedish heart centers. (perform 92% of all bypass surgeries in Sweden) | Consecutive series of 2767 patients with coronary angiography between 5/94 and 1/95 who were considered for coronary revascularization |
| PTCA[12] | 12% (11/95), compared to 9/95 identified by the original panel of NY cardiologists | 27% (26/95), compared to 23/95 identified by the original panel of NY cardiologists | A follow-up to reference 11 was done using a sub-sample of the patients. A panel of Duke University cardiologists reviewed 308 records for appropriateness | |
| Diagnostic testing for Coronary Artery Disease[17] | 3% (7/215) overall | 39% (42/109) overall | Five urban Los Angeles area hospital emergency departments, 2 public, 1 private NFP, 1 university med. ctr., 1 NFP HMO | 356 patients with chest pain not due to myocardial infarction or history of cardiac disease between Oct 94 and Apr 96. Those not receiving ECG during initial eval were excluded |
| Hip Joint Replacement[25] | 8.3% (86/997) overall; 6.7% - 16.3%. for osteoarthritis, 0% - 25.0% for avascular necrosis, 0% for fracture and revision | 32.4% (334/997) overall; 42.3%-50.0% for osteoarthritis, 0% - 50.0% for avascular necrosis, 9.6%-40.0% for fracture, 3.4%-18.9% for revision | 5 large public hospitals (4 university affiliated, 1 community-based) | 997 patients with osteoarthritis, avascular necrosis, hip fracture, or revision who were undergoing HJR between 12/96 and 12/97 |
| Hip and Knee Joint Replacement[21] | High-rate region: 6.1%<br>Low-rate region: 6.4%<br><br>Rated by subspecialists<br>High-rate region: 11.4%<br>Low-rate region: 11.0% | Not evaluated | 7 high-rate region hospitals: 3 university affiliated, 4 community<br>8 low-rate region hospitals: 5 university affiliated, 3 community | 371 patients in the high rate region and 565 in the low rate region with surgery performed between 4/1/92 and 3/31/93 without fracture or other indication, and < 60 years old |

| Procedure | % inappropriate | % uncertain | Source of population | # of patients evaluated |
|---|---|---|---|---|
| Hysterectomy [18] | 16% overall; varied across plans from 10% to 27% | 25% overall | Seven managed care organizations | Random sample of 642 hysterectomies (non-emergency and non-oncological) between 8/1/89 and 7/31/90, among women enrolled in a health plan 2 years prior to surgery |
| Hysterectomy [23] | 70% (367/497); varied from 45% to 100% across diagnoses indicative of hysterectomy | Not evaluated | Nine capitated medical groups in Southern California | 497 women receiving hysterectomy between 8/93 and 7/95 in one of nine capitated medical groups in S. California |
| Laminectomy [19] | 23% | 29% | One Swiss University hospital | 196 patients with surgical treatment for herniated discs |
| Lumbar Discectomy and Spinal Stenosis surgery [20] | 38% (126/328) | Combined with "appropriate" category | Two university neurosurgery departments | 328 consecutive patients undergoing surgery for Lumbar Disc Hernia of Spinal Stenosis at hosp A from 4/92-10/92 and at hosp B from 5/93-9/93. Patients with neoplasms were excluded |
| Upper GI Tract endoscopy [3] | 17% overall; varied from 15% to 19% among sites | 11% overall; varied from 8% to 14% among sites | 5 sites of varying utilization for the 3 procedures selected from Medicare claims submitted by physicians in Arkansas, Colorado, Iowa, Mass., Montana, Penn., S. Carolina, and N. Calif. | Random sample of Medicare beneficiaries for each procedure (claims submitted in 1981) at each site (high, average, and low use geographic areas) |
| Upper GI Tract endoscopy [4] | Varied by county from 0%-25% | Not evaluated | 23 adjacent rural and urban, large and small, counties in one large, populous state | Sampled procedures by Medicare billing codes performed on 614 UGI patients in 1981, aged 65 years and older |

**References**


1. Wong JH, Findlay JM, Surez-Alamazor ME. Regional performance of Carotid Endarterectomy. *Stroke*. 1997; 28:891-898.
2. Matchar DB, et al. Influence of projected complication rates on estimated appropriate use rates for carotid endarterectomy. *HSR*. 1997; 32(3):325-342.
3. Chassin MR, et al. Does inappropriate use explain geographic variations in the use of health care services: A study of three procedures. *JAMA*. 1987;258:2533-2537.
4. Leape LL, et al. Does inappropriate use explain small—area variations in the use of health care services? *JAMA*. 1990;236:669-672.
5. Tobacman JK, et al. Assessment of appropriateness of cataract surgery at ten academic medical centers in 1990. *Ophthalmology*. 1996;103:207-215.
6. Pilpel D, Fraser GM, Kosecoff J, Weitzman S, Brook RH. Regional differences in appropriateness of cholecystectomy prepaid health insurance system. *Public Health Rev*. 1992/93;20:61-74.
7. Froehlich F, et al. Performance of panel-based criteria to evaluate the appropriateness of colonoscopy: A prospective study. *Gastrointest Endosc*. 1998;48:128-36.
8. Noonan SJ, et al. Relationship of anatomic disease to appropriateness ratings of coronary angiography. *Arch Intern Med*. 1995;155:1209-1213.
9. Gray D, Hampton JR, Bernstein SJ, Kosecoff J, Brook RH. Audit of coronary angiography and bypass surgery. *Lancet*. 1990;335:1317-1320.
10. McGlynn EA, et al. Comparison of the appropriateness of coronary angiography and coronary artery bypass graft surgery between Canada and New York State. *JAMA*. 1994;272:934-940.
11. Bernstein SJ, et al. The appropriateness of use of coronary angiography in New York State. *JAMA*. 1993;269:766-769.
12. Leape LL, et al. Effect of variability in the interpretation of coronary angiograms on the appropriateness of use of coronary revascularization procedures. *Am Heart J*. 2000;139:106-113.
13. Leape LL, et al. The appropriateness of Coronary Artery Bypass Graft surgery in academic medical centers. *Ann Intern Med*. 1996; 125:8-18.
14. Leape LL, et al. The appropriateness of use of Coronary Artery Bypass Graft surgery in New York state. *JAMA*. 1993;269:753-760.
15. Ziskind AA, Lauer MA, Bishop G, Vogel A. Assessing the appropriateness of coronary revascularization: The University of Maryland Revascularization Appropriateness Score (RAS) and its comparison to RAND expert panel ratings and American College of Cardiology/American Heart Association guidelines with regard to assigned appropriateness rating and ability to predict outcome. *Clin Cardiol*. 1999;22:67-76.
16. Bernstein SJ, Brorsson B, Aberg T, Emanuelsson H, Brook RH, Werko L. Appropriateness of referral of coronary angiography patients in Sweden. *Heart*. 1999;81:470-477.
17. Carlisle DM, et al. Underuse and overuse of diagnostic testing for coronary artery disease in patients presenting with new-onset chest pain. *Am J Med*. 1999;106:391-398.
18. Bernstein SJ, et al. The appropriateness of hysterectomy: A comparison of care in seven health plans. *JAMA*. 1993; 269: 2398-2402.
19. Porchet F, Vader JP, Larequi-Lauber T, Costanza MC, Burnand B, Dubois RW. The assessment of appropriate indications for laminectomy. *J Bone Joint Surg [Br]*. 1999;81-B:234-239.
20. Larequi-Lauber T, et al. Appropriateness of indications for surgery of Lumbar Disc Hernia and Spinal Stenosis. *Spine*. 1997;22:203-209.
21. Van Walraven C, et al. Appropriateness of primary total hip and knee replacements in regions of Ontario with high and low utilization rates.
22. Karp HR, Flanders WD, Shipp CC, Taylor B, Martin D. Carotid Endarterectomy among Medicare beneficiaries: A statewide evaluation of appropriateness and outcome. *Stroke*. 1998;29:46-52.
23. Broder MS, Kanouse DE, Mittman BS, Bernstein SJ. The appropriateness of recommendations for hysterectomy. *Obstet Gynecol*. 2000; 95:199-205.
24. Wong JH, Lubkey TB, Auarez-Almazor ME, Findlay JM. Improving the appropriateness of Carotid Endarterectomy: Results of a prospective city-wide study. *Stoke*. 199;30:12-15.
25. Quintana JM, Azkarate J, Goenaga JI, Aróstegui I, Beldarrain I, Villar JM. Evaluation of the appropriateness of hip joint replacement techniques. *Intl J Tech Assessment in Health Care*. 2000; 16:165-177.
26. Ferguson JA, Adams TA, Weinberger M. Racial differences in cardiac catheterization use and appropriateness. *Am J Med Sci*. 1998;315:302-306.

Table 2A. Studies of ACSC indicators.

| Study | Payer | Level | Validating variables |
|---|---|---|---|
| **Begley, 1994** | all | area (city/state) | none |
| **Bierman, 2000** | all | person | income (area, all), Medicaid (asthma, diabetes), race (asthma, diabetes), race (asthma, diabetes) |
| **Billings, 1993** | all | area (zip code) | income, race (diagnosis-specific findings in Billings, 1992) |
| **Billings, 1996** | all and Canadian health system | area (zop code) | Income, national health insurance |
| **Bindman, 1995** | all | area (zip code cluster/MSSA) | self-rated access (all, but strongest for asthma, CHF, DM), income, education, uninsured, Medicaid |
| **Blustein, 1998** | Medicare | person | education, income, usual source of care, race, supplemental Medicaid or other insurance, prior physician visits |
| **Casanova, 1995** | Spanish national health system | person | illiteracy (area), unemployment (area), income (area), school enrollment (area), availability of health center, availability of pediatrician |
| **Casanova, 1996** | Spanish national health system | person | family income, paternal/maternal employment, social class, paternal/maternal education, prior physician visit, pediatrician |
| **Claudio, 1999** | all | area | income, % minority, % children under 18, lack of access to preventative health care, poor housing conditions, environmental exposures, genetic suseptibility |
| **Connell, 1984** | Medicare and Medicaid | hospitalization | reside in nursing home, payment source, quality of care (appropriate diagnostic tests, use of IV insulin)_ |
| **Culler, 1998** | Medicare | person | race, education, living alone/with spouse, income, supplemental Medicaid or other insurance, problems obtaining care, health professional shortage area, core SMSA or rural county |
| **Djojonegoro 2000** | all | area (zip code) | income, race (person) |
| **Gadomski, 1998** | Medicaid | person/quarter | enrollment in Maryland Access to Care (Medicaid managed care with gatekeeper), SSI or AFDC recipients, urban residence, race |
| **Gill, 1997** | Medicaid | person | regular source of care, regular care from primary care physician |
| **Gill, 1998** | Medicaid | person | continuity index (function of number of ambulatory visits and unique providers) |
| **Giuffrida, 1999** | English Health Services Area | area | supply of secondary care (physicians, hosp beds), morbidity measures ( limiting l.t. illness, permanently sick), socioeconomic characteristics (pop density, unemploy, no central heating, crowded accomodations, no car, new commonwealth, retired living alone, students, social class 1 & 2, pop mobility), year |
| **Kaestner, 2000** | all | hospital-income category –year | income |
| **Krakauer, 1996** | Medicare | area (HCSA) | race/ethnicity, supplemental Medicaid or HMO coverage, urban residence, income, unemployment, active, generalist, med spec, & surg spec MD/pop ratio, generalists/MD, med spec/MD, surg spec/MD |
| **Laditka, 1999 (JWA)** | Medicare | area (hospital market area) | income |
| **Laditka, 1999 (JHSP)** | Medicaid | area (hospital market area) | Medicaid proportion |
| **McConnochie 1997** | all | area (zip code) | inner city or "other urban" residence (person), late prenatal care, maternal education, uninsured, Medicaid, race, unemployment, lack of telephone/automobile, childhood poverty |
| **Millman, 1993** | all | area | income (all except congenital syphilis, immunization/preventable, dental; gastroenteritis O:E<2) |
| **O' Sullivan 1996** | all | area | geographic region, insurance status |
| **Pappas, 1997** | all | person | race, income (area), uninsured, Medicaid) |

| | | | |
|---|---|---|---|
| **Parchman, 1994** | all | area, (HSA) | FP-GP/pop ratio, gen IM/pop ratio, gen ped/pop ratio, income |
| **Parchman, 1999** | Medicare | person | primary care shortage area, supplemental Medicaid, income, race, education, marital status |
| **Rohrer, 1997** | all | area | geographic region |
| **Schreiber, 1997** | all | area (zip code) | poverty, race, pop density, primary care MD/pop ratio, health professional shortage area, proximity to hospital |
| **Shi, 1999** | all | person | race, income (area), non-MSA, primary care MD, uninsured, Medicare/Medicaid |
| **Shukla, 1997** | all | person | race (all conditions), insurance type (diabetes, hypertension, asthma), localities |
| **Silver, 1997** | Medicare | area (county) | rural regions of Utah (0/4 anemia, 2/4 asthma, 4/4 cellulitis, 2/4 CHF, ¼ DM, ¾ gastric, 2/4 hypertension, 4/4 pneumonia |
| **Solet, 2000** | all | area | income/ poverty status |
| **Taroni, 1997** | Italian health system | area (local health unit) | low overall hospital use |
| **Weissman, 1992** | all | person | uninsured (0/2 rupt appendix, ½ asthma, 2/2 cellulitis, ½ CHF, 2/2 DM, 2/2 gangrene, ½ hypokalemia, ½ immunizable, 2/2 malig. htn. 2/2 pneumonia, ½ pyelo, 2/2 ulcer), Medicaid, race |

Table 3A. Studies of ACSC indicators.

| Study | | Acute Myocardial Infarction | Acute Poliomyelitis | Anemia, Iron deficiency | Anemia, any | Angina | Asthma | Asthma & bronchitis | Cellulitis or abscess | Cellulitis, or other soft tissue infection | Cellulitis, abscess, lymphadenitis | Cellulitis with skin graft | Chest pain | Congestive heart failure | CHF (plus pulmonary edema) | CHF and shock | Congenital syphilis | Convulsions 'A' & 'B' | COPD | Dehydration/ volume depletion | Dental conditions |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Begley, 1994** | **Adult** | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | X | | | | | X | | | | | | | | X | | | X | X | | |
| **Bierman, 2000** | **Adult** | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | X | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | |
| **Billings, 1993** | **Adult** | | | | | X | X | | | | X | X | | | X | | | X | X | X | X |
| | **Peds** | | | X | | X | X | | | | X | X | | | X | | X | X | X | X | X |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | |
| **Billings, 1996** | **Adult** | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | X | | X | X | | | | X | X | | | X | | X | X | X | X | X |
| **Bindman, 1995** | **Adult** | | | | | | X | | | | | | | | X | | | | X | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | |
| **Blustein, 1998** | **Adult** | | | | | X | X | | | | X | X | | | X | | | | X | X | X |
| | **Peds** | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | |
| **Casanova, 1995** | **Adult** | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | X | | X | X | | | | X | | | | X | | X | X | X | X | X |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | |
| **Casanova, 1996** | **Adult** | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | X | | X | X | | | | X | | | | X | | X | X | X | X | X |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | |
| **Claudio, 1999** | **Adult** | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | X | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | |
| **Connell, 1984** | **Adult** | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | |
| **Culler, 1998** | **Adult** | | | | | X | X | | | | X | X | | | X | | | X | X | X | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | |
| **Djojonegoro 2000** | **Adult** | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | X | | X | X | | | | | | | | X | | X | X | X | X | X |
| **Gadomski, 1998** | **Adult** | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | X | | | X | | | | X | | | | | | X | X | | X | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | |
| **Gill, 1997** | **Adult** | | | | | X | X | | | | | | | | X | | | | X | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | X | | | | X | X | | | X | | | | X | X | |
| **Gill, 1998** | **Adult** | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | X | X | | | | X | | | | X | | | X | X | X | X |
| **Guiffrida, 1999** | **Adult** | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | X | | | | | | | | | | | | | | |
| **Kaestner, 2000** | **Adult** | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | X | | X | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | |

| Study | | Acute Myocardial Infarction | Acute Poliomyelitis | Anemia, Iron deficiency | Anemia, any | Angina | Asthma | Asthma & bronchitis | Cellulitis or abscess | Cellulitis, or other soft tissue infection | Cellulitis, abscess, lymphadenitis | Cellulitis with skin graft | Chest pain | Congestive heart failure | CHF (plus pulmonary edema) | CHF and shock | Congenital syphilis | Convulsions 'A' & 'B' | COPD | Dehydration/ volume depletion | Dental conditions |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Krakauer, 1996** | **Adult** | | | | | X | X | | | | X | X | | | X | | | | X | X | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | |
| **Laditka, 1999 (JWA)** | **Adult** | | | | | X | X | | | | X | X | | | X | | | X | X | X | X |
| | **Peds** | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | |
| **Laditka, 1999 (JHSP)** | **Adult** | | | | | X | X | | | | X | X | | | X | | | X | X | X | X |
| | **Peds** | | | X | | X | X | | | | X | X | | | X | | | X | X | X | X |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | |
| **McConnochie 1997** | **Adult** | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | X | | X | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | |
| **Millman, 1993** | **Adult** | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | X | X | | | | X | X | | | X | | X | X | X | X | X |
| **O'Sullivan, 1996** | **Adult** | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | X | | X | X | | | | X | X | | | X | | | X | X | X | X |
| **Pappas, 1997** | **Adult** | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | X | | X | | | | | X | | | | | | | |
| **Parchman, 1994** | **Adult** | | | | | X | X | | | | | | | | X | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | |
| **Parchman, 1999** | **Adult** | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | X | X | | | | X | | | | X | | | X | X | X | |
| **Rohrer, 1997** | **Adult** | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | X | | X | X | | | | X | | | X | | | X | | |
| **Schreiber, 1997** | **Adult** | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | X | | X | X | | | | X | X | | | X | | X | X | X | X | X |
| **Shi, 1999** | **Adult** | | | | | X | X | | | | X | X | | | X | | | X | X | X | X |
| | **Peds** | | | X | | X | X | | | | X | X | | | | | X | X | | X | X |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | |
| **Shukla, 1997** | **Adult** | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | X | | | | X | | | | | | | | | | | | | | |
| **Silver, 1997** | **Adult** | | | | X | | X | | | X | | | | X | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | |
| **Solet, 2000** | **Adult** | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | X | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | |
| **Taroni, 1997** | **Adult** | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | |
| **Weissman, 1992** | **Adult** | | | | | | X | | X | | | | | X | | | | | | | |
| | **Peds** | | | | | | X | | X | | | | | X | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | |

*Removed "diabetes mellitus without mention of complication, not stated as uncontrolled"
# Added benign hypertension and hypertensive renal failure

Table 3A, cont. Studies of ACSC indicators.

| Study | | Diabetes A | Diabetes B | Diabetes C | Diabetes w/ketoacidosis or coma | Diabetes, Uncontrolled | Diabetes with complications | Diabetes, all | Epilepsy or grand mal status | Failure to thrive | Gangrene | Gastroenteritis | Gastrointestinal hemorrhage, upper | High risk/complicated delivery | Hypertension | Hypertension, malignant | Hypoglycemia | Hypokalemia | Immunizable conditions | Immunization-related or preventable | Immunization-preventable pneumonia | Invasive cervical cancer | Kidney/UTI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Begley, 1994** | **Adult** | | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | X | | X | | | | | | | | X | | | X | | | |
| **Bierman, 2000** | **Adult** | | | | | X | | | | | | | | | | | | | | | X | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | | |
| **Billings, 1993** | **Adult** | X | X | X | | | | | X | | | X | | | X | | X | | | X | | | X |
| | **Peds** | X | X | X | | | | | X | X | | X | | | X | | X | | | X | | | X |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | | |
| **Billings, 1996** | **Adult** | | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | X | X | X | | | | | X | X | | X | | | X | | X | | | X | | | X |
| **Bindman, 1995** | **Adult** | X | X | X | | | | | | | | | | | X | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | | |
| **Blustein, 1998** | **Adult** | X | X | X | | | | | X | | | X | | | X | | X | | | X | | | X |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | | |
| **Casanova, 1995** | **Adult** | | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | X | X | X | | | | | X | X | | X | | | X | | X | | | X | | | X |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | | |
| **Casanova, 1996** | **Adult** | | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | X | X | X | | | | | X | X | | X | | | X | | X | | | X | | | X |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | | |
| **Claudio, 1999** | **Adult** | | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | | |
| **Connell, 1984** | **Adult** | | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | X | | | | | | | | | | | | | | | |
| **Culler, 1998** | **Adult** | X | X | X | | | | | X | | | X | | | X | | X | | | | | | X |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | | |
| **Djojonegoro 2000** | **Adult** | | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | X | X | X | | | | | X | | | X | | | X | | X | | | X | | | X |
| **Gadomski, 1998** | **Adult** | | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | X | X | X | | | | | X | X | | X | | | | | X | | | X | | | X |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | | |
| **Gill, 1997** | **Adult** | X | X | X | | | | | | | | | | | X | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | X | X | X | | | | | | | | X | | | X | | X | | | | | | |
| **Gill, 1998** | **Adult** | | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | X | X | X | | | | | X | | | X | | | X | | X | | | | | | |
| **Guiffrida, 1999** | **Adult** | | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | | |
| **Kaestner, 2000** | **Adult** | | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | X | | X | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | | |

** Separate indicators for pregnancy (646.6+590.1) and others (590.1); removed chronic or NOS pyelonephritis

Table 3A, cont. Studies of ACSC indicators.

| Study | | Diabetes A | Diabetes B | Diabetes C | Diabetes w/ketoacidosis or coma | Diabetes, Uncontrolled | Diabetes with complications | Diabetes, all | Epilepsy or grand mal status | Failure to thrive | Gangrene | Gastroenteritis | Gastrointestinal hemorrhage, upper | High risk/complicated delivery | Hypertension | Hypertension, malignant | Hypoglycemia | Hypokalemia | Immunizable conditions | Immunization-related or preventable | Immunization-preventable pneumonia | Invasive cervical cancer | Kidney/UTI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Krakauer, 1996** | **Adult** | X | X | X | | | | | | | | X | | | X | | X | | | | | | X |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | | |
| **Laditka, 1999 (JWA)** | **Adult** | X | X | X | | | | | X | | | X | | | X | | X | | | X | | | X |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | | |
| **Laditka, 1999 (JHSP)** | **Adult** | X | X | X | | | | | X | | | X | | X | X | | X | | | X | | | X |
| | **Peds** | X | X | X | | | | | X | X | | X | | | X | | X | | | X | | | X |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | | |
| **McConnochie 1997** | **Adult** | | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | X | | X | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | | |
| **Millman, 1993** | **Adult** | | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | X | X | X | | | | | X | | | X | | | X | | X | | | X | | | X |
| **O'Sullivan, 1996** | **Adult** | | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | X | X | X | | | | | X | | | X | | | X | | X | | | X | | | X |
| **Pappas, 1997** | **Adult** | | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | X | | | | | | X | | | | | X | | X | X | | | | |
| **Parchman, 1994** | **Adult** | X | X | X | | | | | | | | | | X | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | | |
| **Parchman, 1999** | **Adult** | | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | X | X | X | | | | | X | | | X | | | X | | X | | | | | | X |
| **Rohrer, 1997** | **Adult** | | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | X | | | | | | | X | | | | | | | | |
| **Schreiber, 1997** | **Adult** | | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | X | X | X | | | | | X | X | | X | | | X | | X | | | X | | | X |
| **Shi, 1999** | **Adult** | X | X | X | | | | | X | | | X | | | X | | X | | | X | | | X |
| | **Peds** | X | X | X | | | | | X | X | | X | | | | | | | | X | | | X |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | | |
| **Shukla, 1997** | **Adult** | | | | | | | | | | | | | | | | | | | | | X | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | X | | | | | | | | | X | | | | | | | | |
| **Silver, 1997** | **Adult** | | | | | | | X* | | | | | X | | X# | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | | |
| **Solet, 2000** | **Adult** | | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | | |
| **Taroni, 1997** | **Adult** | | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | | |
| **Weissman, 1992** | **Adult** | | | | X | | | | | | X | | | | | X | | X | X | | | | |
| | **Peds** | | | | X | | | | | | X | | | | | X | | X | X | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | | |

** Includes other respiratory TB (012)

Table 3A, cont. Studies of ACSC indicators.

| Study | | Low infant birth weight | Malignant neoplasm of female breast | Measlres | Mumps | Nutritional Deficiency | Otitis Media | Pelvic Inflammatory Disease | Perforated or Bleeding Ulcer | Pneumonia, bacterial (outpatient treatable) | Pneumonia, bacterial (any) | Pneumonia or pleurisy | Pneumonia, Bronchitis, Bronchiolotis, Pharyngitis | Pneumonia, sinsusitis | Pyelonephritis | Respiratory infections (including pulmonary TB) | Rheumatic fever | Ruptured appendix | Severe ENT infection | Stroke with hypertension | TB, other | TB, pulmonary |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Begley, 1994** | **Adult**<br>**Peds**<br>**Adult & Peds** | | | | | | | | | | | | | X | X**<br>* | | | | | X | | |
| **Bierman, 2000** | **Adult**<br>**Peds**<br>**Adult & Peds** | | | | | | | | | | | | | | | | | | | | | |
| **Billings, 1993** | **Adult**<br>**Peds**<br>**Adult & Peds** | | | | | X<br>X | | X<br>X | | X<br>X | | | | | | | | | X<br>X | | X<br>X | X<br>X |
| **Billings, 1996** | **Adult**<br>**Peds**<br>**Adult & Peds** | | | | | X | | X | | X | | | | | | | | | X | | X | X |
| **Bindman, 1995** | **Adult**<br>**Peds**<br>**Adult & Peds** | | | | | | | | | | | | | | | | | | | | | |
| **Blustein, 1998** | **Adult**<br>**Peds**<br>**Adult & Peds** | | | | | X | | X | | | | | | | | | | | X | | X | X |
| **Casanova, 1995** | **Adult**<br>**Peds**<br>**Adult & Peds** | | | | | X | | X | | X | | | | | | | | | X | | X | X |
| **Casanova, 1996** | **Adult**<br>**Peds**<br>**Adult & Peds** | | | | | X | | | | X | | | | | | | | | X | | X | X |
| **Claudio, 1999** | | | | | | | | | | | | | | | | | | | | | | |
| **Connell, 1984** | **Adult**<br>**Peds**<br>**Adult & Peds** | | | | | | | | | | | | | | | | | | | | | |
| **Culler, 1998** | **Adult**<br>**Peds**<br>**Adult & Peds** | | | | | | | | | X | | | | | | | | | X | | | |
| **Djojonegoro 2000** | **Adult**<br>**Peds**<br>**Adult & Peds** | | | | | X | | X | | X | | | | | | | | | X | X | X | X |
| **Gadomski, 1998** | **Adult**<br>**Peds**<br>**Adult & Peds** | | | | | X | | | | X | | | | | | | | | X | | X | X |
| **Gill, 1997** | **Adult**<br>**Peds**<br>**Adult & Peds** | | | | | | | | | | X | | | | | | | | X | | | |
| **Gill, 1998** | **Adult**<br>**Peds**<br>**Adult & Peds** | | | | | | | | | | X | | | | | | | | X | | | |
| **Guiffrida, 1999** | **Adult**<br>**Peds**<br>**Adult & Peds** | | | | | | | | | | | | | | | | | | | | | |
| **Kaestner, 2000** | **Adult**<br>**Peds**<br>**Adult & Peds** | | | | | | | | | | | | | | | | | | | | | |

** Separate indicators for pregnancy (646.6+590.1) and others (590.1); removed chronic or NOS pyelonephritis

Table 3A, cont. Studies of ACSC indicators.

| Study | | Low infant birth weight | Malignant neoplasm of female breast | Measlres | Mumps | Nutritional Deficiency | Otitis Media | Pelvic Inflammatory Disease | Perforated or Bleeding Ulcer | Pneumonia, bacterial (outpatient treatable) | Pneumonia, bacterial (any) | Pneumonia, or pleurisy | Pneumonia, Bronchitis, Bronchiolotis, Pharyngitis | Pneumonia, sinsusitis | Pyelonephritis | Respiratory infections (including pulmonary TB) | Rheumatic fever | Ruptured appendix | Severe ENT infection | Stroke with hypertension | TB, other | TB, pulmonary |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Krakauer, 1996** | **Adult** | | | | | | | | | X | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | |
| **Laditka, 1999 (JWA)** | **Adult** | | | | | X | | X | | X | | | | | | | | | X | | X | X |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | |
| **Laditka, 1999 (JHSP)** | **Adult** | | | | | X | | X | | X | | | | | | | | | X | | X | X |
| | **Peds** | | | | | X | | X | | X | | | | | | | | | X | | X | X |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | |
| **McConnochie 1997** | **Adult** | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | |
| **Millman, 1993** | **Adult** | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | X | | | | | | | | | X | | | |
| **O'Sullivan, 1996** | **Adult** | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | X | | X | | X | | | | | | | | | X | | | X |
| **Pappas, 1997** | **Adult** | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | X | | X | | | | X | | | X | | | | |
| **Parchman, 1994** | **Adult** | | | | | | | | | X | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | X | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | |
| **Parchman, 1999** | **Adult** | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | X | | | | | | | | | X | | | |
| **Rohrer, 1997** | **Adult** | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | X | | | | | X | | | | X | | | | | | |
| **Schreiber, 1997** | **Adult** | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | X | | X | | X | | | | | | | | | X | | X | X |
| **Shi, 1999** | **Adult** | | | | | X | | X | | X | | | | | | | | | X | | X | X |
| | **Peds** | | | | | X | | | | X | | | | | | | | | X | | X | X |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | |
| **Shukla, 1997** | **Adult** | | X | | | | | | | | | | | | | | | | | | | X |
| | **Peds** | X | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | X | X | | | | | | | | | | | | X | | | | | X* |
| **Silver, 1997** | **Adult** | | | | | | | | | | | | X | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | |
| **Solet, 2000** | **Adult** | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | |
| **Taroni, 1997** | **Adult** | | | | | | | | | | | | | | | | | | | | | |
| | **Peds** | | | | | | | | | | | | | | | | | | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | |
| **Weissman, 1992** | **Adult** | | | | | | | | X | | X | | | | X | | | X | | | | |
| | **Peds** | | | | | | | | X | | X | | | | X | | | X | | | | |
| | **Adult & Peds** | | | | | | | | | | | | | | | | | | | | | |

* Includes other respiratory TB (01)

Table 4A. Studies of ACSC Indicators- Pediatric Avoidable Hospitalizations

| Study | Acute Ottisis Media | Acute pyelonephritis | Asthma | Bronchitis | Burns | Congenital Lower Bowel Obstruction*** | Diabetic ketoacidosis | Gastroenteritis | Hypernatremic*** | Immunizable Diseases | Jaundice*** | Mastoiditis | Nausea and Vomiting | Obstuctive Cardiac Defects*** | PID | Upper Respiratory Infection | UTI | Viral Meningitis | Viral Syndrome | Volume depletion*** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Gadomski, 1998** | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |

*** Infant readmission within 2 weeks of DOB

Table 5A. Studies of ACSC Indicators- Discretionary Hospitalization of Infants

| Study | Croup/Tracheitis | Failure to Thrive | Fever, acute unexplained | Gastroenteritis/ Dehydration | Gastrointestinal disorders, other | Lower Respiratory Illness | Nonspecific signs and symptoms | Seizure w/o intractactable mentioned | UTI | Viral Meningitis |
|---|---|---|---|---|---|---|---|---|---|---|
| **Kaestner, 2000** | X | X | X | X | X | X | X | X | X | X |
| **McConnochie, 1997** | X | X | X | X | X | X | X | X | X | X |

Table 6A. Studies of ACSC Indicators- Late Hospitalization

| Study | Appendicitis | Diverticular disease | Hernia, External | Peptic Ulcer | Pneumonia, bacterial | Uterine Fibroma |
|---|---|---|---|---|---|---|
| **Taroni, 1997** | X | X | X | X | X | X |

## References


1 Begley, C, Slater CH, Engel MJ, Reynolds TF. Avoidable hospitalization and socio-economic status in Galveston County, Texas. *J Comm Health*. 1994;19:377-387.

2 Bierman AS, Steiner C, Friedman B, Fillmore CM, Clancy CM. Ambulatory Care Sensitive Conditions: A viable indicator of access to quality primary care. Abstract. Available at http://www.ahsr.org/1999/abstracts/bierman.htm. Accessed: 5/18/00.

3 Billings J, Anderson GM, Newman LS. Recent findings on preventable hospitalizations. *Health Affairs*. 1996;15:239-249.

4 Billings J, Zeitel L, Lukomnik J, Carey TS, Blank AE, Newman L. Impact of socioeconomic status on hospital use in New York City. *Health Affairs*. 1993; Spring:162-173

5 Bindman AB, Grumbach K, Osmond D, et al. Preventable hospitalizations and access to health care. *JAMA*. 1995;274:305-311.

6 Blustein J, Hanson K, Shea S. Preventable hospitalizations and socioeconomic status. *Health Affairs*. 199817:177-189.

7 Cassanova C, Colomer C, Starfield B. Pediatric hospitalization due to ambulatory care-sensitive conditions in Valencia (Spain). *Intl J Quality in Health Care*. 1996;8:51-59.

8 Cassanova C, Starfield B. Hospitalizations of children and access to primary care: A cross-national comparison. *Intl J of Health Services*. 1995;25:283-294.

9 Claudio L, Tulton L, Doucette J, Landrigan PJ. Socioeconomic factors and asthma hospitalization rates in New York City. *J Asthma*. 1999;36:343-350.

10 Connell FA, Blide LA, Hanken MA. Clinical correlates of small area variations in population-based admission rates for diabetes. *Med Care*. 1984;22:939-949.

11 Culler SD, Parchman ML, Przybylski M. Factors related to potentially preventable hospitalizations among the elderly. *Med Care*. 1998;36:804-817.

12 Djojonegoro BM, Williams AF. Area Income as a predictor of preventable hospitalizations in the Harris County Hospital District, Houston. *Tex Med*. 2000;96:58-62.

13 Epstein AJ. The role of the medical market in preventable hospitalizations [abstract]. *Abstract Book/Association for Health Services Research*, 1998:15:316-7.

14 Gadomski A, Jenkins P, Nichols M. Impact of a Medicaid provider and preventive care on pediatric hospitalization. *Peds*. 1998;101(3):E1.

15 Gill JM, Mainous AG. The role of provider continuity in preventing hospitalizations. *Arch Fam Med*. 1998;7:352-357.

16 Gill JM. Can hospitalizations be avoided by having a regular source of care? *Fam Med*. 1997;29:166-71.

17 Giuffrida A, Gravelle H, Roland M. Measuring quality of care with routine data: avoiding confusion between performance indicators and health outcomes. *BMJ*. 1999;319:94-98.

18 Halterman, JS, Aligne A, Auinger P, McBride JT, Szilagyi PG. Inadequate therapy for asthma among children in the United States. *Peds*. 2000;105:272-276.

19 Kaestner R, Racine A, Joyce T. Did recent expansions in Medicaid narrow socioeconomic differences in hospitalizations rates of infants. *Med Care*. 2000;38:195-206.

20 Krakauer H, Jacoby I, Millman M, Lukomnik JE. Physician impact on hospital admission and on mortality rates in the Medicare population. *HSR*. 1996;31:191-211.

21 Laditka SB, Johnston JM. Preventable hospitalization and avoidable maternity outcomes: Implications for access to health services for Medicaid recipients. *J Heath & Social Policy*. 1999;11:41-55.

22 Laditka SB, Laditka JN. Geographic variation in preventable hospitalization of Older Women and Men: Implications for access to primary health care. *J Women & Aging*. 1999;11:43-56.

23 McConnochie KM, Roghmann KJ, Liptak GS. Socioeconomic variation in discretionary and mandatory hospitalization of infants: An ecologic analysis. *Peds*. 1997;99:774-784.

24 O'Sullivan MJ, Volicer B. Preventable Hospitalizations: A tool for planning and marketing ambulatory health care services. *J Ambulatory Care Manage*. 1996;19:84-95.

25 Ordoñez GA, Phelan PD, Olinsky A, Robertson CF. Preventable factors in hospital admissions for asthma. *Arch Dis Child*. 1998;78:143-147.

26 Pappas G, Hadden WC, Kozak LJ, Fisher GF. Potentially avoidable hospitalizations: Inequalities in rates between US socioeconomic groups. *Am J Pub Health*. 1997;87:811-816.

27 Parchman ML, Culler S. Primary care physicians and avoidable hospitalization. *J Fam Prac*. 1994;39:123-128.
28 Parchman ML, Culler SD. Preventable hospitalizations in primary care shortage areas: An analysis of vulnerable Medicare Beneficiaries. *Arch Fam Med*. 1999;8:487-491.
29 Rohrer JE, Vaughan M. Monitoring health care system performance in Iowa. *Health Srvs Manage Research*. 1997;10:107-112.
30 Schreiber S, Zielinski T. The meaning of ambulatory care sensitive admissions: Urban and Rural Perspectives. *J Rural Health*. 1997;13:276-284.
31 Shi L, Samuels ME, Pease M, Bailey WP, Corley EH. Patient characteristics associated with hospitalizations for ambulatory care sensitive conditions in South Carolina. *S Med J*. 1999;92:989-998.
32 Shukla RK, Pestian JP. Small area analysis of ambulatory care sensitive conditions in Virginia [abstract]. *Abstract Book/Association for Health Services Research*. 1997;14:90.
33 Silver MP, Babitz ME, Magill MK. Ambulatory care sensitive rates in the aged Medicare population in Utah, 1990 to 1994: A rural-urban comparison. *J Rural Health*. 1997;13:285-294.
34 Solet D, Krieger J, Stout J, Lui L. Childhood asthma hospitalizations – King County, Washington, 1987-1998. *CDC MMWR Weekly*. www.cdc.gov/mmwr/preview/mmwrthml. 2000; 49:929-933. Last accessed 10/19/00.
35 Taroni F, Repetto F, Louis DZ, Moro ML, Yuen EJ, Gonnella JS. Variation in hospital use and avoidable patient morbidity. *J Health Services Research and Policy*. 1997;2:217-222.
36 Weissman JS, Gatsonix C, Epstein AM. Rates of avoidable hospitalization by insurance status in Massachusetts and Maryland. *JAMA*. 1992;268:2388-2394.

# APPENDIX 9

# List of Acronyms

List of acronyms and meanings used in the evidence report.

## Appendix 9. List of Acronyms Used

| | |
|---|---|
| AAA | Abdominal aortic aneurysm |
| ACG | Ambulatory Care Groups |
| ACSC | Ambulatory care sensitive conditions |
| ADE | Adverse drug events |
| ADL | Activities of Daily Living |
| AHCPR | Agency for Health Care Policy and Research |
| AHRQ | Agency for Healthcare Research and Quality |
| AMI | Acute myocardial infarction |
| APR-DRGs | All Patient Refined – Diagnosis Related Groups (developed by 3M™ Corporation) |
| CABG | Coronary artery bypass graft |
| CAD | Coronary artery disease |
| CCP | Cooperative Cardiovascular Project |
| CE | Carotid endarterectomy |
| CHF | Congestive heart failure |
| CONQUEST | Computerized Needs-oriented Quality Measurement Evaluation System |
| COPD | Chronic obstructive pulmonary disease |
| CPI | Consumer Price Index |
| CSI | Comprehensive Severity Index |
| C-section | Cesarean section |
| DEMPAQ | Develop and Evaluate Methods to Promote Ambulatory Care Quality (HCFA) |
| DNR | Do not resuscitate |
| DRG | Diagnosis Related Groups, developed by HCFA |
| DVT | Deep vein thrombosis |
| EPC | Evidence-Based Practice Center |
| GI | Gastrointestinal |
| HCFA | Health Care Financing Administration |
| HCUP | Healthcare Cost and Utilization Project |
| HEDIS | The Health Plan Employer Data and Information Set |
| HP 2000 or 2010 | Healthy People 2000 or 2010, developed by the Dept. of Health and Human Services |
| ICD-9-CM | The International Classification of Diseases, ninth revision, clinical modification |
| IRB | Internal review board |
| IT | Information technology |
| JCAHO | Joint Commission on Accreditation of Healthcare Organizations |
| LC | Laparoscopic cholecystectomy |
| MEDPAC | Medicare Payment Advisory Commission |
| MeSH | Medical Subject Headings (developed by National Library of Medicine for Medline) |
| MDC | Major Diagnostic Category |
| MSA | Metropolitan Statistical Area |
| MSX | Multivariate Signal Extraction |
| NCQA | National Committee on Quality Assurance |
| NIS | Nationwide Inpatient Sample |
| NLHI | National Library of Healthcare Indicators |
| OLS | Ordinary Least Squares (regression) |
| OMD | Optimal minimal distance |
| ORYX™ | Trademark name used by JCAHO for a set of hospital performance measures |
| OSHPD | Office of Statewide Health Planning and Development, State of California |
| PID | Pelvic Inflammatory Disease |
| PIP-DCG | Principal Inpatient Diagnostic Cost Groups, developed by HCFA |
| PPV | Positive predictive value |
| PTCA | Percutaneous transluminal coronary angioplasty |
| QI | Quality Indicators |
| RCT | Randomized controlled trial |
| SAS™ | formerly, Statistical Analysis System (distributed by the SAS Corporation) |
| SEER | Surveillance, Epidemiology, and End Results |

| | |
|---|---|
| SPSS™ | formerly, Statistical Package for Social Scientists (distributed by the SPSS Corporation) |
| UB-92 | Uniform Billing form 92, developed by HCFA |
| UCSF | University of California, San Francisco |
| UR | Uterine rupture |
| UTI | Urinary tract infection |
| VAR | Vector autoregression structure |
| VBAC | Vaginal birth after cesarean section |