

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, bilateral 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.