

CMS
Centers for Medicare & Medicaid Services

Special Study

**Identification and Synthesis of
Components Essential to
Achieving "High Performer"
Status in Various Provider Types**

Task 2: High Performer "State of the Art" Report

February 2004

HSAG
HEALTH SERVICES
ADVISORY GROUP

1600 East Northern Avenue, Suite 100 • Phoenix, AZ 85020
Phone 602.264.6382 • Fax 602.241.0757

Executive Summary 1-1

Performance Measurement..... 1-2
 Key Clinical Indicators of Quality Performance..... 1-3
 Key Quality Characteristics Associated With High Performers..... 1-3
 Summary of Findings 1-4
 Findings Regarding Key Quality Characteristics Associated With High Performers 1-4
 Findings Regarding Key Clinical Indicators of Quality Performance 1-4
 Conclusions..... 1-5

Overview and Methodology 2-1

Background 2-1
 The Changing Health Care Environment 2-1
 Search Protocols 2-1
 Initial Review 2-2
 Final Protocol and Relevant Findings 2-3
 Performance Measurement 2-3
 Key Clinical Indicators of Quality Performance 2-6
 Key Quality Characteristics Associated With High Performers 2-10
 Critique of the Literature Pertinent to High Performance 2-14
 Conclusion: Four Principles 2-17

Tables 3-1

References and Endnotes 4-1

Appendices

Appendix A—Study Protocol for Defining High Performing Health Care Providers..... A-1
 I. Introduction and Background A-1
 II. Purpose of the Study..... A-1
 III. Review Questions A-1
 A. Performance Measurement..... A-1
 B. Key Clinical Indicators of Quality Performance A-2
 IV. Review Protocol..... A-2
 A. Literature Review..... A-2
 B. Non-Literature Sources to Identify Key Quality Characteristics of Performance Measures A-5
 C. Information Synthesis..... A-5
Appendix B—Results of MeSH Search on Performance Methods..... B-1
Appendix C—Results of MeSH Search on Key Clinical Indicators of Quality Performance C-1

Tables

Table 3-1—Literature Describing Methods Used to Identify High Performers, by Health Care Setting 3-2

Table 3-2—Literature Evaluating Various Statistical Profiling Methods 3-6

Table 3-3—7th SoW Clinical Conditions Used in Measuring Institutional or Practitioner Performance in Hospitals..... 3-8

Table 3-4—7th SoW Clinical Conditions Used in Measuring Institutional or Practitioner Performance,
by Other Settings 3-9

Table 3-5—Quality Improvement Dimensions Affecting Quality Performance Addressed in Literature and
Organizational Material 3-10

Appendices

Appendix A—Study Protocol for Defining High Performing Healthcare Providers A-1

Figure A-1—Data Abstraction Tool A-6

Table A-1—Selection Criteria and Inclusion and Exclusion Criteria for Literature Review A-3

Table A-2—Selection Criteria and Inclusion and Exclusion Criteria for Abstract Review A-4

Appendix B—Results of MeSH Search on Performance Methods..... B-1

Figure B-1—MeSH Search Performance Methods..... B-1

Table B-1—Results of MeSH Search on Performance Methods B-2

Appendix C—Results of MeSH Search on Key Clinical Indicators of Quality Performance C-1

Figure C-1—Key Clinical Indicators of Quality Performance C-1

Table C-1—Results of MeSH Search on Key Quality Characteristics of Performance C-2

1. Executive Summary

Under contract with the Centers for Medicare & Medicaid Services (CMS), Health Services Advisory Group, Inc. (HSAG) is conducting a Special Study on *Identification and Synthesis of Components Essential to Achieving “High Performer” Status in Various Provider Types*. This review of published literature and other pertinent information is the second task in this Special Study. The purpose of the literature review is to define high performance, and synthesize information that has been used to identify and characterize hospitals that perform exceptionally well on Seventh Scope of Work (7th SoW)[‡] and proposed Eighth Scope of Work (8th SoW) process indicators that are part of the management of three conditions: (1) pneumonia, (2) heart failure, and (3) acute myocardial infarction (AMI). Synthesis and dissemination of the results can facilitate efforts to improve quality in all hospitals and to provide tools that Quality Improvement Organizations (QIOs) can use in working with hospitals. The importance of this project has increased with the passage on December 8, 2003, of the Medicare Prescription Drug, Improvement, and Modernization Act of 2003 (Medicare Modernization Act), which provides financial incentives to hospitals to voluntarily provide CMS with quality data to be used for public reporting purposes.

In addition to the medical literature, the review included information from Arizona hospitals participating in the Three State Pilot (3SP)[†] program, as well as descriptions of awards for (or recognition of) high performance used by other QIOs and other organizations interested in quality standardization. An initial review of 317 abstracts from articles published between 2000 and 2003 failed to identify any that addressed questions regarding what defines “high performer” or what constitutes high performance with regard to the three conditions from the 7th SoW. As the review proceeded, HSAG developed a working definition of a high performing hospital. A high performing hospital is one that:

- ◆ achieves a performance score (proportion of services provided) of at least 90 percent, on at least one indicator, when there are two indicators.
- ◆ achieves a performance score of at least 90 percent, on at least two indicators, where there are three or more indicators.
- ◆ achieves a 95th percentile score on the “hard” indicator(s).*
- ◆ achieves a 95th percentile score on the “easy” indicator(s).*
- ◆ achieves a 90th percentile score, on all indicators between the easy indicator and the hard indicator, when there are more than two indicators.
- ◆ sustains these standards for two consecutive six-month or four consecutive quarterly periods.

[‡] Scope of Work (SoW) refers to three-year contract periods in which quality improvement organizations conduct activities on behalf of the Centers for Medicare & Medicaid Services (CMS). The 7th SoW covers the period November 1, 2002, through October 31, 2005.

[†] The Three State Pilot (3SP) refers to a three-state pilot project launched in December 2002 by the U.S. Department of Health and Human Services, Centers for Medicare & Medicaid Services. The quality improvement organizations in Arizona, Maryland, and New York were asked to test the most effective ways to communicate with consumers about hospital quality of care.

* For purposes of this review, a “hard” indicator is defined as one that few, if any, hospitals do well on, and an “easy” indicator is defined as one on which the vast majority of hospitals do well.

Based on the results of the initial review, the search protocol was revised to focus on two specific areas, each with its own research question(s):

1. Performance measurement:

- What statistical methods, based on existing quality indicators, have been used for profiling and ranking institutions, and for classifying their performance?
- Do these methods discriminate among institutions?
- Are they meaningful to decision makers?

2. Key clinical indicators of quality performance:

- To what extent have clinical conditions and related performance measures from the current 7th SoW and proposed 8th SoW been the subjects of published research identifying high performance in hospitals, nursing homes, home care agencies, and physician offices?

Performance Measurement

Articles addressing performance measurement were of three types:

- ◆ articles using datasets to profile high performing institutions, using a variety of profiling methods
- ◆ articles evaluating performance measures used for profiling
- ◆ theoretical articles discussing the application of statistical techniques in profiling

Profiling techniques included:

- ◆ Bayesian hierarchical modeling^{41,42}
- ◆ data envelopment analysis (DEA)
- ◆ stochastic frontier analysis (SFA)^{43,44}
- ◆ hierarchical generalized linear models (HGLM)⁴⁵
- ◆ Achievable Benchmarks of Care (ABC)³⁹
- ◆ standard descriptive techniques applied to cross-sectional data⁴⁶⁻⁴⁸

Of these, DEA, SFA, and probably Bayesian methods, as well, were judged too complex for purposes of this review. ABC benchmarking was, by far, the easiest to use and understand; it also lends itself to modification, permitting its use with smaller sample sizes. One study found that there was poor association between Joint Commission on Accreditation of Healthcare Organizations (JCAHO) measures and a set of measures, largely financial in nature, derived from Medicare data, thereby pointing out the fact that high performance in one area does not assure high performance in other areas. Measures used in profiling hospitals, nursing homes, and physician providers were predominantly outcomes measures, such as mortality rates,^{41,50-54} complication rates,⁵²⁻⁵⁴ and patient satisfaction.^{48,55}

Recent methodological literature emphasizes the advantages of process over outcomes indicators. Outcomes are most appropriate when emphasis is on *how* something is done rather than on *whether* it is done. The major limitation on process indicators has to do with sample size and the dependence of results on clusters of patients attended by a sample of different physicians that was not randomized. Hierarchical logistic regression techniques are recommended to minimize systemic variation, reduce the number of outliers, and improve separation between and within cluster variability.^{15,17,59}

Key Clinical Indicators of Quality Performance

Key clinical indicators of quality performance of hospitals relating to the management of AMI and heart failure were overwhelmingly related to mortality; however, one article pointed out that mortality was not useful in distinguishing between hospitals. For pneumonia, timeliness of blood culture and antibiotic administration were the most commonly cited process indicators. Pneumonia management was cited as an index of emergency room (ER) quality, as a measure of improvement in *small* hospitals, and as a measure of statewide public health practice. Process-of-care indicators for heart disease and pneumonia have been used successfully to measure improvement in hospital care. In the areas of nursing home care, home care, and physician care, the issues in which researchers show the greatest interest are pressure sores in nursing home patients and diabetes screening in physician office patients. Performance measures were used to explore compliance with best practices for these conditions and to calculate risk-adjusted pressure ulcer rates to profile nursing homes.

Representatives from rural hospitals in Arizona believed their hospitals would never be viewed as institutions delivering quality care if they were separated out from larger hospitals when quality was measured. They believed that an important characteristic of a high performing hospital was that its performance exceeded that of other hospitals, regardless of hospital size or location. All hospital representatives thought that reporting data to consumers was best accomplished through the use of composite scores, supplemented by graphics, which showed the individual indicators making up the composite. These hospital representatives were concerned that achieving the level of improvement sufficient for voluntary public reporting would require a methodology with face validity. They failed to agree on the feasibility of weighting clinical indicators, but they did not disagree with the idea that doing well on an indicator that hospitals had trouble with was worth more as a quality measure than doing well on an indicator on which the vast majority of hospitals did well.

Key Quality Characteristics Associated With High Performers

To complete the task of identifying key quality characteristics, literature and non-literature quality-oriented sources of information were reviewed. These sources included a convenience sample of published articles; nonpublished manuscripts; business-oriented quality assessment groups, including the International Organization for Standardization (ISO) program—the National Institute of Standards and Technology’s Malcolm Baldrige Award; and an accrediting body—JCAHO. Additionally, nine QIOs that had a formal recognition and awards program for health care organizations were queried to ascertain their measurement strategies of performance. HSAG then compared the performance dimensions deemed important by these organizations to further clarify

characteristics of high performers to be included in the qualitative phase (Task 5) of this Special Study.

Summary of Findings

Findings Regarding Key Quality Characteristics Associated With High Performers

A number of organizational characteristics and quality improvement practices are thought to be associated with institutional high performance as evidenced by the respective assessment frameworks established by three of the most well-known and accepted health care quality management and accreditation programs: Baldrige, ISO, and JCAHO. Although these programs differ in their approaches to measuring quality, they focus on many of the same dimensions of organizational quality improvement: cultural; structural; strategic; leadership; systems; customer and market; measurement, analysis, and knowledge management; human resources; and business results.

Together, these nine dimensions of organizational quality improvement have been studied to varying degrees in health services research over the past three decades. The studies have addressed the question of which organizational characteristics and quality improvement practices are associated with institutional high performance. Studies have been directed largely toward understanding relationships between organizational structure attributes and cost, service mix, and staffing. Until we gain an understanding of the linkages between structure, process, and outcome, we will continue to be hampered in our efforts to improve quality.¹⁰⁵ What is becoming more clear is that what matters is not whether a hospital possesses the knowledge of quality improvement (QI) principles, tools, and techniques, but the extent to which progress is made in implementing them.¹⁰¹ Complexity theory suggests that understanding can only be achieved when we study the organizational attributes in concert, rather than individually.¹¹⁷

Findings Regarding Key Clinical Indicators of Quality Performance

The measurement of high performance is very much a work in progress. While growing demand for professional and public accountability has led to a proliferation of health care “report cards” in one form or another,^{1,2} the technical aspect of such reporting remains a challenge. Outstanding issues include the problematic use of outcomes measures due to problems of inadequate and often competing risk adjustment schemas; sampling problems of process measures including the clustering of hospital patients with the same condition by attending physicians; need for more flexibility in the use of summary or composite measures; and the desirability for, but continuing lag in, the use of electronic records. Other difficult problems include controversy over weighting of indicators; lack of a method of determining the reliability of performance profiles; the need to incorporate patient satisfaction into performance measures; and the incorporation of more relevant structural measures, such as nurse staffing levels. The public reporting of data calls attention to the importance of methodological issues, even though the literature does not support the hope that consumers can or will use performance data in choosing providers.

Conclusions

Neither the published medical literature nor material from quality rating organizations directly addressed the Special Study's focus on hospitals that are high performing in their management of the three conditions from the 7th SoW. However, with the concurrence of the review findings, the hospital meetings, and conversations with consultants and among the collaborators to this Special Study, it has become possible to identify four principles that will guide HSAG in proceeding to a final definition of high performer and the development of the algorithm to identify them. These principles are:

1. Hospitals that provide the same service should be held to the same standard, regardless of size or geographic location.
2. Performing well on an indicator on which others perform poorly (a "hard" indicator), while performing as well as any on an indicator on which others do well (an "easy" indicator), is basic to a definition of high performance.
3. Performing well on "hard" indicators is more important than performing well on "easy" indicators.
4. High performance must incorporate sustained excellence over time.

The information synthesized from the current review has reaffirmed CMS' position that there is an underlying need within the health care system to design a scientifically sound and credible structure by which high performing health care providers can be defined, identified, and studied. Furthermore, the information synthesized from the current review has established a high degree of feasibility to design such a structure. The goal of this special study is to provide direction and opportunity to move the health care field one step closer to understanding the mechanisms by which organizational characteristics mediate quality performance. Such an understanding would afford additional opportunity for QIOs to provide assistance for improving health care quality to all stakeholders by replicating successful models.

Background

HSAG is conducting this Special Study, *Identification and Synthesis of Components Essential to Achieving “High Performer” Status in Various Provider Types*, under contract with CMS. The second task of this project is to describe the “state of the art” about high performance/high performers in the health care settings covered by Task 1 of the 7th SoW: hospitals, nursing homes, home health agencies, and physician offices. To this end, this report draws on published and non-published sources to summarize what is known about measuring high performance on clinical conditions from the 7th SoW and from the proposed 8th SoW in these health care settings, and what methods have been used to measure and compare institutional performance.

Information on measuring and discriminating organizational performance will be used to refine a working definition of high performance. This definition will be operationalized and tested using available publicly reported process of care indicators for the four provider settings, starting with hospitals. Additionally, the definition will be used to develop an algorithm to identify high performing health care organizations, beginning with hospitals.

The Changing Health Care Environment

At the beginning of this review, it became increasingly clear that the Medicare Modernization Act would be enacted, including an all-but-mandatory public reporting requirement for hospitals. This will add hospitals to the list of virtually all other providers (except individual physicians) whose data are publicly reported by CMS. Because this Special Study focuses primarily on hospitals, and because public reporting is a totally new activity for most hospitals, HSAG wanted to define and characterize high hospital performance in a way that would provide individual hospitals with an understanding of how they could maximize their own performance. This aim engenders a strong bias toward developing an algorithm that is relatively simple, easily understood by hospital personnel, and conducive to inter-hospital comparison.

Search Protocols

This review was not intended to be an exhaustive, academically oriented survey of the literature. The intent was rather to identify the recent work of health care leaders in measuring and defining performance that might help characterize organizational high performers/high performance in selected health care settings for selected conditions from the 7th SoW and the proposed 8th SoW. Developing the study protocol for the literature review was a two-step process. The initial literature review following the original study protocol was not useful for defining high performance. In consultation with the CMS project team, study questions, search goals, and search terms were refined and a more targeted second review undertaken.

Initial Review

HSAG’s initial literature review study protocol for defining high performer/high performance queried PubMed (an online journal citation service of the National Library of Medicine [NLM]) using various combinations of medical subject heading (MeSH) terms such as hospitals, nursing homes, physicians, home care, managed care, quality of care, quality indicators/health care, quality assurance/health care, organization and administration, pneumonia, congestive heart failure, myocardial infarction, and surgical wound infection. In that review, we restricted the search to articles published in English between the years 2000 and 2003.

While 317 abstracts were retrieved and examined, few of these abstracts helped in addressing the question: “How do you define ‘high performer’?” For example, the 186 abstracts on managed care and 104 on administration and organization yielded little insight into the study purpose. Researchers addressed many dimensions of performance (e.g., access, availability, medical errors, best practices with feedback to providers, outcomes measures) generating a complex view of potential performance measures. At that point, we developed a working definition of high performance, limited to three clinical conditions from the 7th SoW hospital public reporting initiative: (1) pneumonia, (2) heart failure, and (3) AMI.

A high performing hospital for a given condition is one that:

- ◆ achieves a performance score (proportion of services provided) of at least 90 percent, on at least one indicator, when there are two indicators.
- ◆ achieves a performance score of at least 90 percent, on at least two indicators, where there are three or more indicators.
- ◆ achieves a 95th percentile score on the “hard” indicator(s).*
- ◆ achieves a 95th percentile score on the “easy” indicator(s).*
- ◆ achieves a 90th percentile score, on all indicators between the easy indicator and the hard indicator, when there are more than two indicators.
- ◆ sustains these standards for two consecutive six-month or four consecutive quarterly periods.

In order to refine this working definition, we initiated a second study protocol (Appendix A). Specific changes from the initial protocol included: the elimination of the managed care, administration, and organization quality of care literature; greater emphasis on the statistical methods used to measure high performance; and the merging of two research questions targeted at the 7th and 8th SoW health care settings and clinical conditions into a more focused question (see Appendix A, section III-A).

* For purposes of this review, a “hard” indicator is defined as one that few if any hospitals do well on, and an “easy” indicator is defined as one on which the vast majority of hospitals do well.

Final Protocol and Relevant Findings

The literature and non-literature review findings are organized according to the final study protocol's two subject areas: (1) performance measurement, and (2) key clinical indicators of quality performance in hospitals, nursing homes, home care, and physician offices.

Performance Measurement

The literature search for this subject area focused on these questions:

- ◆ What statistical methods, based on existing quality indicators, have been used for profiling and ranking institutions and for classifying their performance?
- ◆ Do these methods discriminate between high and non-high performing institutions?
- ◆ Are these statistical methods meaningful to decision makers?

The literature search on performance measurement targeted methods articles published in English between 1997 and 2003. PubMed was searched for combinations of the key words: hospital, nursing homes, home care, physicians, statistics and numerical data, statistical models, benchmarking, quality indicators/health care, and health services research. Overall, 101 citations were obtained and 55 abstracts were examined for appropriateness for inclusion in this review (Appendix B). Of these, 13 articles met the inclusion criteria that were used to identify studies that reported on relevant statistical methods used to measure and profile institutions. In addition, 14 citations were included in the review that were identified from article references or in consultation with experts. Articles were excluded that primarily focused on economic or fiscal analyses, and on utilization or manpower studies. A total of 27 articles were reviewed for the study and are included in the review database.

This section reviews the literature and discusses the statistical methods and formats used in profiling hospitals, nursing homes, and physicians.

Review of the literature on statistical methods identified three types of articles: (1) those using datasets to profile high performing institutions, using a variety of profiling methods; (2) articles evaluating performance measures used for profiling; and (3) theoretical articles discussing the application of statistical techniques in profiling. The literature related to identifying high performing health care providers and the use of performance measures is found in Table 3-1. Table 3-2 summarizes the literature on statistical approaches and issues related to benchmarking and profiling techniques.

Profiling Methods Used to Identify High Performers

The profiling methods used with datasets were Bayesian hierarchical modeling,^{41,42} data envelopment analysis (DEA), stochastic frontier analysis (SFA),^{43,44} hierarchical generalized linear

models (HGLM),⁴⁵ Achievable Benchmarks of Care (ABC),³⁹ and standard descriptive techniques applied to cross-sectional data.⁴⁶⁻⁴⁸ (See Table 3-2.)

Berlowitz and colleagues⁴² reported that Bayesian hierarchical modeling allowed for the development of better statistical models for profiling nursing homes, by accounting for additional sources of variation within and among institutions. He also reported Bayesian models as complex to develop and requiring sophisticated model-checking methods to ensure that the chosen model fits the data. Using an HGLM over time, Bronskill et al.⁴⁵ found that longitudinal profiles were a useful supplement to more commonly used cross-sectional analyses. One particular advantage is that longitudinal techniques, using multiple years of data, help to reduce sampling variability.

Using DEA analysis to identify best performers, O’Neal et al.⁴⁴ found that the researchers needed adjustments to the output variables to make them appropriate for the DEA model, thus likely underestimating the number of efficient models. Similarly, Jacobs⁴³ reported DEA methods less sensitive to efficiency scores than SFA analysis.

Relying on basic descriptive techniques, computation of ABCs was the technically easiest method to apply. Kiefe³⁹ reported on the application of ABCs to two sets of data and described the refinements to this benchmark methodology, using a Bayesian estimator technique and a “minimum sufficient denominator” formula to deal with the issue of small sample size.

One study⁴⁹ evaluated the content validity, reliability and sensitivity, and the rank order comparisons of a set of nine utilization and clinical performance measures derived from Medicare reports and found that the most useful measures for comparing hospitals were related to asset turnover, cash flow, mortality, complications, and cost per case. These measures were subsequently used²⁴ to compare a Medicare data set with the overall evaluation scores from a JCAHO dataset. Only two of the Medicare dataset measures had a significant association with JCAHO overall evaluation scores, and one of these associations was in an unexpected direction. The authors suggest that different criteria and measurement methods from those used by JCAHO are needed for evaluating hospital performance.

Profiling Measures

Measures used in profiling hospitals, nursing homes, and physician providers were predominantly outcomes measures such as mortality rates,^{41,50-54} complication rates,⁵²⁻⁵⁴ and patient satisfaction.^{48, 55} Selected clinical process of care measures,^{46,51,56} clinical procedures,^{47,52} and utilization/fiscal measures^{48,53-55} were also used in studies distinguishing performance among health care providers.

Theoretical Literature

There were eight articles that addressed various methodological issues in the use of statistical profiling methods (Table 3-2). This section summarizes the major recommendations of researchers for assessing health care quality.

Using routine data for quality assessment purposes is attractive to researchers because of their availability, because they represent a rich source of information about large numbers of patients, and because they are less costly and difficult to obtain.⁵⁸ Two articles^{58,59} address the challenges of using outcomes data for quality measures. These include difficulties with the validity and reliability of measures; the demands of case mix adjustment; chance variability; and poor quality data. All may lead to artifacts in measuring quality using outcomes data.^{58,59} With the exception of data quality, the use of process measures is less susceptible to these measurement issues. Mant⁵⁹ offers six recommendations with respect to the selection of process or outcomes measures for monitoring the quality of care:

1. Process measures that are valid, relevant, and practical should be used in preference to outcomes measures.
2. Consider outcomes measures if it is likely that *how* something is done is as important as *what* is done, or if the complication rate of a procedure is of particular importance.
3. Complete power calculations before setting up an outcomes study.
4. Consider the costs for an outcomes monitoring system. Data must be accurate, complete, and of sufficient quantity to make adjustments for case mix.
5. Perform a cost-benefit analysis for the outcomes monitoring project, accounting for provider fear, accusation, and defensiveness.
6. Interpret results considering an estimate of the plausible impact of quality of care on outcome, power of the study, and the extent to which selection bias may account for observed differences.

Shahian and colleagues⁶⁰ discuss problems of random interprovider variation in determining outliers, provider sample size differences, and the use of aggregate results from patient-level prediction models to assess cardiac surgeon performance. To manage these concerns about reporting and comparing mortality rates of cardiac surgeons, the researchers recommend the use of hierarchical generalized linear regression models. These models result in less systematic variation, broader confidence intervals for regression parameters, and fewer outliers.

Process measures are more sensitive than outcomes measures in detecting differences in quality.⁵⁹ They are also actionable and have face validity. Unfortunately, they are only applicable to a selected subset of the study population.⁶¹ This limitation makes consideration of sample size an issue in process-based quality studies. Normand and Zou⁶¹ focus attention on determining sample sizes for measuring and comparing the quality of care delivered to cohorts of patients by different health care providers. The data used for inference involve observations on units grouped within clusters, such as patients treated within hospitals, resulting in selection bias from non-randomization. Focusing on sample size determination for process-based measures, the researchers demonstrate that hierarchical

sampling models permit separation of between-and-within cluster variability, thereby improving inferences for individual clusters.

Allison¹¹ posed four similar issues that need to be addressed with the use of inferential statistics in quality measurement. These included the appropriate use of confidence intervals, use of inferential statistics for ranking providers, the need for risk adjustment with outcomes measures, and calculation of aggregate quality measures and their corresponding measures of statistical significance. Finally, and most importantly, Allison et al.¹¹ remind us that feedback to providers of benchmark performance levels on process indicators is best used as a tool to improve performance, not as a tool to ascertain whether a provider is a statistical outlier.

In summary, recent methodological literature emphasizes the advantages of process over outcomes indicators. Outcomes are most appropriate when emphasis is on *how* something is done rather than on *whether* it is done. The major limitation on process indicators has to do with sample size and the dependence of results on clusters of patients attended by a sample of different physicians that was not randomized. Hierarchical logistic regression techniques are recommended to minimize systemic variation, reduce the number of outliers, and improve separation of between-and-within cluster variability.

Key Clinical Indicators of Quality Performance

The literature search for this subject area focused on the following question: To what extent have clinical conditions and related performance measures from the current 7th SoW and proposed 8th SoW been the subjects of published research identifying high performance in hospitals, nursing homes, home care, and physician offices?

The literature search to identify the key clinical indicators of quality in measuring performance in hospitals, nursing homes, home care, and physician offices targeted articles published in English between the years 2000 and 2003, and involved health care providers in the United States, the United Kingdom, or Canada. Combinations of MeSH major topics, keywords, and words in abstracts were queried in PubMed. Search words included the four provider types, conditions from the 7th SoW (see Tables 3-3 and 3-4) and proposed 8th SoW, as well as the terms “quality indicators/health care” and “quality assurance/health care.” For this aspect of the study, we also searched for citations related to the MeSH major topics of Medicare and professional review organization by the four provider types. A total of 184 citations were obtained and abstracts were examined for appropriateness for inclusion in this review. Abstracts excluded from consideration were those that did not address the appropriate setting, population, or condition. The total number of abstracts meeting inclusion criteria for review was 36 (Appendix C, Figure C-1).

To complete the task of identifying key quality characteristics, literature and non-literature quality-oriented sources of information were reviewed. These sources included a convenience sample of published articles; nonpublished manuscripts; business-oriented standardization groups, including the International Organization for Standardization (ISO) programs⁹⁸ and the National Institute of

Standards and Technology's Malcolm Baldrige Award⁹⁷ program; and accrediting bodies such as JCAHO.¹⁰⁷ Additionally, nine QIOs that had a formal recognition and awards program¹⁰⁸⁻¹¹⁶ for health care organizations were queried to ascertain their performance measurement strategies. HSAG then compared the performance dimensions deemed important by these QIOs to further clarify characteristics of high performance to be included in the qualitative study.

Finally, HSAG drew upon discussions with Arizona hospitals related to improved performance and how to best present data. These discussions took place over a 13-month period beginning in November 2002 with Arizona's entry into the 3SP designed to recruit hospitals to voluntarily report data on 10 process indicators related to the management of the three 7th SoW conditions: (1) pneumonia, (2) heart failure, and (3) AMI. Seventy-five percent of all Arizona hospitals have volunteered for the 3SP, including 95 percent of urban hospitals that accounted for 98 percent of all hospital discharges in 2001. HSAG meets with hospitals every month, and about 30 hospitals are represented at each meeting. Unless there is a guest speaker, separate monthly meetings are held in Phoenix and Tucson to minimize travel time and improve attendance. With passage of the Medicare Modernization Act, HSAG expects to reach nearly 100 percent participation over the next few months.

Literature—Hospitals

After combining the hospital and "Medicare/PRO" MeSH searches and eliminating duplicates, there were 18 articles included in the hospital abstract review (Appendix C). We identified an additional article from consultation with experts, yielding a total of 19 abstracts reviewed. Of these, acute myocardial infarction (AMI) received the most attention from researchers, followed by heart failure and pneumonia. There were no citations found on surgical wound infection (Table 3-3).

The primary AMI and heart failure performance measure for distinguishing among hospitals was mortality rates measured as standardized mortality ratios,^{41,65} risk-standardized mortality ratios,⁶⁶ and unadjusted^{67,68} and risk-adjusted⁶⁹ mortality rates. With one exception,⁶⁷ mortality rates were found not to be useful for distinguishing among hospitals.

AMI indicators from the 7th SoW that received the most attention from researchers (Table 3-3) were the use of aspirin and beta blockers during hospitalization, and timely reperfusion.^{56,66,67,70,71} Only two studies were found that used any 7th SoW indicators to measure quality of inpatient heart failure care.^{56,72}

AMI and heart failure studies also described the association of hospital ownership with patient outcomes,⁷³ the effect of provider and hospital characteristics on guideline implementation,^{71,74} the effects of quality improvement (QI) projects on adherence to best practices,^{68,70} and institutional resources and clinical approaches that lead to improvements in AMI care.⁷⁵

Pneumonia process-of-care measures were used in QI projects to measure improvement in care in small hospitals;⁷⁶ in emergency departments;⁷⁷ at the state level, to improve process-of-care

performance in hospitals;⁷⁸ and to identify patient and hospital characteristics associated with variability in pneumonia care performance.⁷⁹ With respect to profiling, development of post-operative pneumonia was used in combination with mortality rates to enhance detection of outlier hospitals,⁸⁰ and publicly reported indicators of pneumonia care were used at the state⁸¹ and national⁵⁶ levels. The 7th SoW process of care indicators reported most frequently for pneumonia were timing of blood cultures and timeliness of antibiotic administration. (See Table 3-3 for a complete list of quality indicators).

Literature Review—Other Providers

Searches were also carried out on 7th and proposed 8th SoW conditions for providers in other settings: nursing homes, home care, and physician offices. Of the 34 citations found for nursing homes, 14 met exclusionary criteria (Appendix C). Among the 10 studies addressing the target nursing home conditions, 7 measured aspects of pressure ulcers, 2 measured the use of physical restraints, and 1 addressed patient confusion (Table 3-4). Five of the 7 pressure ulcer studies used the Minimum Data Set (MDS) indicators on pressure ulcer. These studies were designed to: derive and validate a risk adjustment model for pressure ulcer development;^{82,118} monitor nursing homes using risk-adjusted rates of pressure ulcer development;⁸³ measure reliability and feasibility of a pressure ulcer scoring protocol;⁸⁴ and determine whether nursing homes scoring in the extreme quartiles of the pressure ulcer indicator provided different pressure ulcer care.⁸⁵ Berlowitz⁸² found that nursing homes differed significantly in their expected rates of pressure ulcer development and that outlier identification also differed depending on whether unadjusted or risk-adjusted performance is evaluated. Bates-Jensen et al.,⁸⁵ not using risk adjustment techniques, were unable to demonstrate that documented pressure ulcer care was related to pressure ulcer prevalence rates in nursing homes. A sixth study examined the effects of QI implementation and organizational culture on pressure ulcer care⁸⁶ and found no significant association between QI implementation and either compliance with pressure ulcer guidelines or the rate of pressure ulcer development. The final study on pressure ulcers⁸⁷ was a case study that found significant differences between bedside nurses and managers on their choice of the major clinical indicators of high quality nursing home care, with nurses selecting pressure ulcers and managers selecting hydration management.

One of the two studies (Table 3-4) on the use of physical restraints in nursing homes examined the characteristics of nursing homes and nursing home residents that were associated with deficiency citations for physical restraint use,⁸⁸ and found low registered nurse staffing levels, high Medicaid census, and high numbers of residents that required assistance with activities of daily living (ADL) were associated with state survey and certification citations of poor quality care. The second study described one organization's strategies to target physical restraint reduction.⁸⁹ In the single study on confusion in nursing home residents (Table 3-4), a valid and reliable observational instrument was developed for measuring the physical environment of institutional settings for persons with dementia.⁹⁰

There were six citations found for the targeted conditions in home care, but only one of these met inclusion criteria (Appendix C). This study (Table 3-4)⁹¹ identified a number of demographic, clinical, and functional factors that predicted repeat hospitalizations.

In the physician PubMed search, there were 4 of 33 citations that met inclusion criteria (Appendix C). Three studies (Table 3-4) examined physician compliance in testing glycosylated hemoglobin and lipid profiles and ordering eye exams.⁹²⁻⁹⁴ These studies found that: nurse practitioners were more likely than physicians to document many diabetes care processes, including glycosylated hemoglobin;⁹³ improvement in adherence to guidelines was difficult to sustain over time;⁹² and that physician inattention, patient noncompliance, systems issues and guideline exceptions based on comorbidities were common reasons physicians cited for non-adherence to guidelines. One study (Table 3-3) explored physician approaches to breast cancer screening⁹⁵ and found that only 42 percent of physicians agreed with 75 percent or more of best practice screening recommendations.

In summary, the conditions targeted by the 7th SoW indicators of care on which health care researchers have focused were: (1) pneumonia, (2) heart failure, and (3) AMI. There were disappointing results using mortality measures for distinguishing hospital performance. The process-of-care indicators for heart disease and pneumonia have been used successfully to measure improvement in care. Among nursing homes, home care, and physicians, the conditions with the greatest interest among researchers are pressure sores in nursing home patients and diabetes screening in physician offices. Performance measures were used to explore compliance with best practices for these conditions and to risk-adjust pressure ulcer rates to profile nursing homes.

Arizona Hospitals

Discussions with Arizona hospitals were more diffuse and do not fall readily into specific categories. Nevertheless, they were invaluable and contributed at least as much useful information as any other source. What is described here is either the result of direct discussion of a topic or was implicit in general discussions about hospital efforts to improve on 7th SoW and 3SP indicators. Rural hospital representatives were greatly concerned about “bypasses,” the tendency of small town residents to ignore the local hospital and travel to a big city for hospitalization. They believed that separating hospitals into urban versus rural categories promotes this practice, and it was they who first agreed that small and large hospitals should be held to the same standard. The urban hospital representatives agreed with this view.

Weighting of indicators came up in several meetings. Those in favor of it, however, recognized the difficulty of getting general agreement on how the weights should be distributed. It was from this discussion that the principles emerged that while all indicators should be unweighted in terms of clinical importance, it would be appropriate to regard high performance on indicators that most hospitals struggled with as more indicative of high performance than results on less challenging indicators. The hospital representatives were unanimous in the view that presentation of data on the <http://www.medicare.gov/> Web site should represent composite scores derived from individual indicators available in the form of a graphic, such as a radar chart.* There was concern about the need for face validity. They thought it would be difficult to “sell” a reporting program if the physician staff could not understand the methodology.

* Radar charts, sometimes referred to as star or spider graphs, have multiple axes along which data can be plotted. They are primarily used as a means of comparison between data sets.

Key Quality Characteristics Associated With High Performers

Quality Improvement Dimensions Synthesized from Published and Non-published Sources

A number of organizational characteristics and quality improvement practices are elements thought to be associated with institutional high performance. For the purpose of discussion in this report, these elements have been grouped into categories, which are referred to as “dimensions” of quality improvement. Combinations and modifications of assessment frameworks drawn from sources included in the current review comprised the following nine QI dimensions,^{97,98,101,107} which encompass and address those elements thought to be associated with high performers:

1. *Cultural Dimension*—Encompasses the underlying beliefs, values, norms, and behaviors of the organization that support, or serve as barriers to, quality improvement.
2. *Structural Dimension*—Encompasses the specific internal groups and reporting mechanisms used to lead the organization’s quality improvement activities and establish accountability.
3. *Strategic Dimension*—Encompasses the extent to which the organization’s key strategic priorities and overall business plan are driving its quality improvement agenda.
4. *Leadership Dimension*—Addresses how senior executives guide the organization and how the organization addresses good citizenship and its responsibilities to the public at large.
5. *Systems Dimension*—Addresses aspects of how key productivity, delivery, and support processes are designed, managed, and improved.
6. *Customer and Market Dimension*—Addresses how the organization determines requirements and expectations of customers and markets, builds relationships with customers, and acquires, satisfies, and retains customers.
7. *Measurement, Analysis, and Knowledge Management Dimension*—Addresses the management, effective use, analysis, and improvement of data and information to support key organizational processes and the organization’s performance management system.
8. *Human Resources Dimension*—Addresses how the organization enables its workforce to develop its full potential and how the work force is aligned with the organization’s objectives.
9. *Business Results Dimension*—Addresses the organization’s performance and improvement in its key business areas and how the organization performs in relation to its competitors.

Elements included in these dimensions have been studied to varying degrees in health services research, and used by well-known organizations to assess health care provider quality performance.

Well-Known Organizations Assessing Health Care Provider Performance Using Elements of the Nine Quality Improvement Dimensions

Three of the most well-known organizations using elements of these quality improvement dimensions to assess health care provider quality performance are the International Organization of Standardization (ISO), the National Institute of Standards and Technology (NIST) Baldrige National

Quality Program (Baldrige Program), and the Joint Commission on Accreditation of Healthcare Organizations (JCAHO). The purpose of accreditation programs, such as JACHO, is to assess organizations against a standard of performance that is judged to be acceptable. They have traditionally set a minimum threshold level for quality management and measured against it.

Effective January 2004, JCAHO will implement an initiative to progressively sharpen the focus of its assessment process for hospitals, ambulatory care, behavioral health care, home care, laboratory, and long term care providers,¹⁰⁷ to expand on elements of the nine quality improvement dimensions listed above. The new initiative targets “Critical Focus Areas (CFAs)” that are known to significantly impact patient safety and quality of care. The changes from the previous assessment approach are primarily deletions, consolidations, or clarifications of existing standards. The changes were implemented in an effort to make the assessment more relevant for physicians and to increase the involvement of the direct care-giving staff.¹⁰⁷

Like JCAHO, the ISO sets standards against which organizations’ quality performance is assessed. ISO has drafted an International Standard, titled *Quality Management Systems (QMS)—Guidelines for Process Improvement in Healthcare Organizations*. QMS is based on eight principles that reflect elements of the quality improvement dimensions listed above. The QMS principles address the following eight areas: customer focus, leadership, involvement of people, process approach, system approach to management, continual improvement, factual approach to decision making, and mutually beneficial supplier relationships.

Similarly, the purpose of the Baldrige National Quality Program is to facilitate and guide an organization to continuously improve quality. The difference with respect to the Baldrige Program is that it encourages organizations to set and achieve increasingly tough “stretch goals” for itself. The program’s “Health Care Criteria for Performance Excellence” have three important roles: (1) to help improve organizational performance practices, capabilities, and results; (2) to facilitate communication and sharing of best practice information; and (3) to serve as a working tool for understanding and managing performance and for guiding organizational planning and opportunities for learning. The “Core Values and Concepts” that underlie the Health Care Criteria include the following: visionary leadership, patient-focused excellence, organizational and personal learning, valuing staff and partners, agility, focus on the future, managing for innovation, management by fact, social responsibility and community health, focus on results and creating value, and systems perspective.⁹⁷

Health Services Research Addressing Relationships Between Quality Performance and Elements of the Nine Quality Improvement Dimensions

Elements included in the nine quality improvement dimensions have also been studied to varying degrees in health services research over the past three decades. The studies have addressed the question of which organizational characteristics and quality improvement practices are associated with institutional high performance.

In an early study, Shortell set the stage by proposing that high performance is defined in terms of the ability to provide high-quality cost-effective care, to innovate, to grow, and to enjoy an excellent reputation in the community and among one's professional staff and employees. He further proposed that high performers share the following cultural characteristics: they "stretch themselves, maximize learning, take risks, exhibit transforming leadership, exercise a bias for action, create a chemistry among top managers, exhibit a loose coherence, exhibit a well-defined culture, and reflect a basic spirituality."⁹⁹

Results from the continuing stream of research in the Cultural Dimension have suggested other characteristics that may influence organizational performance in continuous quality improvement. A culture that demonstrates a strong management commitment to continuous quality improvement (CQI), respects professional values, and allows for autonomy and flexibility is conducive to CQI. Also, a culture that encourages open, accurate, and timely communications consisting of clear, concrete CQI objectives supports CQI efforts.^{97-99,102,104,106}

Studies addressing issues in the Strategic Dimension emphasize the importance of tying the institution's quality improvement efforts to strategic priorities and aligning organizational incentives with improvement objectives.^{97,98,101,103,107}

In the Customer and Market Dimension, the literature emphasizes the need to focus on the customers, both internal and external, and include them in decision-making. Specialized nursing units have been suggested as a means to better meet the customers' (patients') needs.^{97,98,101,102,105,107}

In relation to the Leadership Dimension, there is a sense that a transformational leadership style, one that uses a "manager as developer" model and demonstrates a bias for action, results in a more effective CQI program. It is also thought that modeling care improvement on the part of senior executives and setting high standards with clear expectations while encouraging initiative will facilitate organizational CQI efforts.^{97-99,101,103,104}

Literature addressing the Human Resources Dimension stresses the use of both intrinsic and extrinsic motivators. Implementation of a comprehensive approach toward learning and knowledge building is used to empower staff, and alignment of CQI objectives with the compensation system is used to reward staff.^{97,98,100-102,105,107}

In the Measurement, Analysis, and Knowledge Management Dimension literature, the argument was made that at a minimum it is necessary to establish a reliable flow of useful information. Data must be gathered regularly, regardless of the level of technology. Ideally, a strong information systems technology should be a high priority for the organization. In terms of knowledge management, importance was placed on maximizing learning at the individual and group levels within the organization. Decentralized decision-making and education and training in the use of CQI tools and techniques were regarded as essential.^{97-99, 102,103,105,107}

In relation to the Structural Dimension, cross-sectional CQI teams and an organization-wide quality council with physician involvement was seen as instrumental to successful CQI efforts. Designated units primarily responsible for quality improvement were also reported as advantageous.^{97,98,101,106,107}

QIO Formal Performance Award and Recognition Criteria Using Elements of the Nine Quality Improvement Dimensions

QIOs that had formal recognition and awards programs for health care organizations were queried to ascertain the criteria used in their programs. A verbal request was made during the November 2003 conference call of the American Health Quality Association's Health Care Quality Improvement Program (HCQIP) Network, immediately following an overview of the High Performers Special Study. That verbal solicitation was followed by an e-mail to all QIO HCQIP Network representatives asking for information on QIOs' awards and recognition program criteria, and a written reminder was included in the minutes of the November conference call sent to the same individuals. In response, criteria for awards and recognition programs were received from nine QIOs sponsoring such programs.¹⁰⁸⁻¹¹⁶ HSAG classified each state's criteria by the nine quality improvement dimensions. The classification results summarized below were used to further clarify characteristics of high performers to be included in the qualitative phase (Task 5) of this Special Study.

The Measurement, Analysis, and Knowledge Management Dimension had the greatest number of criteria reported, and represented all nine states. The nine QIOs reported using one or more of 19 criteria. The criteria addressed the following: methods and frequency of data collection and reporting; percentage of performance improvement on a variety of 7th SoW quality indicators; demonstrated sustainability of improvement; reduction in failure rate; absolute performance rate; collaborator rankings by performance rates; minimum number of indicators demonstrating improvement; and minimum number of clinical topic areas showing improvement.

The Systems Dimension had the second greatest number of criteria reported, and represented the second greatest number of states reporting. Seven of the nine states used one or more of 10 criteria. The criteria addressed the following: description of process or systems change; description of results following interventions; utilization of multi-faceted approach; completion of process improvement plans and intervention plans; and description of barriers and challenges encountered.

The Leadership Dimension was in third place, based on both the number of criteria reported and number of states represented. The five reporting QIOs used five award and recognition criteria. The criteria focused primarily on demonstration of innovation and initiative, with one criterion addressing education to the community at large.^{109,111,112,114,116}

The Cultural Dimension placed fourth overall in the number of QIOs represented and number of criteria established for recognition or award. Three states used a combination of six criteria that related to various aspects of collaboration and sharing.^{111,115,116}

The Strategic Dimension also represented three states, but only five criteria were used—all of which addressed levels of participation in QIO projects.^{111,115}

The Structural Dimension was addressed by only two of the nine states with only two criteria, which both dealt with implementation of CQI teams.

None of the states reported basing award and recognition criteria on the Human Resources or Business Results Dimensions.

Critique of the Literature Pertinent to High Performance

The substance of this critique is taken largely from published editorial comments by Jeroan Allison, MD, MS, and subsequent discussions with Dr. Allison and his colleague, Norm Weissman, PhD.^{7,11,67}

The measurement of high performance is very much a work in progress. While growing demand for professional and public accountability has led to a proliferation of health care “report cards” in one form or another,^{1,2} technical aspects of such reporting remains a challenge. Recent peer-reviewed literature places less emphasis on the theoretical and technical aspects of these reports of quality measurement than on interventions to improve quality. McGlynn et al.³ suggest that current reporting systems could be improved by the following:

1. Increased attention to overall care by using several quality measures per condition, rather than the usual single condition or single measure per condition, and by focusing on a broad scope of diagnoses that encompass most of the reasons for hospitalization.

Such breadth discourages shunting disproportionate resources toward improving only specific quality measures at the expense of other conditions.^{4,5} A broad measurement base also decreases the possibility of “gaming,” or manipulating patient cohorts to achieve artificially high results. For example, using data from three large health maintenance organizations, Hofer⁶ demonstrated that mean glycosylated hemoglobin levels for poor-performing physicians could be dramatically improved by removing only a few selected patients from the profile. This susceptibility to manipulation remained even after extensive case-mix adjustment.

2. Flexibility sufficient to allow multiple levels of summary measures. Summary measures may be more meaningful to consumers than an overwhelming array of individual measures.

McGlynn et al.³ speculate that examination at the system level, by condition and by individual measure, may suggest complementary, yet distinct, approaches to improvement. Flexible aggregation exposes poor performance on individual measures or dimensions that otherwise might remain hidden within composite measures. Conversely, a review of component measures may disclose the cause of a low aggregate score.

3. Greater efficiency of medical record abstraction by using a single data set for patients with different combinations of conditions.

An electronic chart abstraction module, customized to the individual record during manual abstraction, guides data collection. This approach allows potentially meaningful differences to be detected with relatively small sample sizes. The resulting efficiency decreases the expense of chart abstraction, although manual abstraction will become less relevant with increasing adoption of electronic medical record systems.⁷

A consensus statement of the American Medical Association (AMA), JCAHO, and the National Committee for Quality Assurance (NCQA) summarized the desirable attributes of performance measures.⁸ This statement posits that important quality measures should address target conditions which exert a large population impact and which manifest demonstrable variation in care, thus revealing potential for improvement. Useful measures engender recommendations that are evidence based, achievable, and under the direct control of the physician or health care organization. Furthermore, measures with sound design (1) require a properly specified numerator, denominator, and sampling methodology; (2) possess high accuracy, validity (including ease of interpretation), and reliability; (3) allow appropriate risk adjustment; (4) depend on data that are economically and logistically feasible to acquire; (5) allow independent verification through auditing techniques; (6) guard against dishonest manipulation; and (7) maintain confidentiality.

There are important technical difficulties that require additional attention.^{9,10} For example, although less important in “rapid cycle feedback” for internal QI, the concept of precision demands close scrutiny when quality measures are used for public reporting.^{11,12} As the unit of reporting becomes more specific, small and uneven sample sizes may decrease reliability.¹³ Multiple comparisons and failure to account for the clustering of patients by physician, within health plans, may artificially inflate statistical significance.^{14,15} Few measurement systems describe the reliability, or internal consistency, of performance profiles.¹⁶ Finally, weighting individual measures to build aggregate scores is complex, and not always necessary. Weighting is most straightforward when relevant objective risks and benefits are clearly quantified.¹⁷ Consideration of patient preferences, a necessary step in ultimately defining appropriateness, introduces additional complexity to weighting.¹⁸

Public reporting to assist consumers with choosing among health providers is also fraught with difficulty. Among these problems is the observation that public reporting has only moderate to no influence on provider choice.^{19,20} Those with a higher socioeconomic position usually have more options when performance of their current plan is inadequate. Low literacy creates difficulty in interpreting complex reports. Furthermore, variability within and between plans makes the choice more difficult.²¹ Such variability implies that a consumer could not receive top-quality from any single plan.

Difficulty arises as well from the inability to accurately characterize Donabedian’s classic quality triad of structure, process, and outcomes.²² Accreditation of health care institutions, which relies heavily on structural measures, does not assure high quality performance.²³ Many structural measures have no relationship or even a counter-intuitive relationship with outcomes.²⁴ At the same time, there is a growing realization that certain structural measures, such as nurse staffing levels, are closely related to outcomes.²⁵ Three new and more relevant structural measures—computerized order entry, use of high volume providers, and provider board certification—have been proposed by The Leapfrog Group.²⁶

Outcomes, or the “end results of health care,” extend beyond mortality to broader measures that are meaningful to the patient, such as quality of life and satisfaction.²⁷ For fair comparison across providers, institutions, and delivery systems, outcomes measures require risk adjustment. Many risk-adjustment schemes have been designed, but different methods lead to different conclusions, and no method comprehensively addresses the multiple sources of variability affecting outcomes.²⁸ In addition, intermediate outcomes measures, such as level of lipid control or blood pressure, do not consider unmeasured but crucial factors such as appropriateness of patient use of medication.

Process measures, although more sensitive to poor quality than outcomes, are relevant only when they lead to desired outcomes.¹⁷ That a clinical trial demonstrates that a process conveys a proven benefit is no guarantee that measuring that process and reporting performance will improve outcomes.²⁹ Appropriate process measures depend upon a precisely defined denominator of “ideal candidates,” or those with clear indications and no contraindications for the treatment.³⁰ Given the imperfection of current data systems and the inability to fully capture nuances of clinical reasoning, defining the ideal patient is difficult. Shekelle^{31,32} notes that even the RAND appropriateness method, which entails a rigorous approach to combining expert opinion and scientific evidence, does not produce recommendations suitable for patient-level decisions.

Societal and personal expectations of the health care system provide the context for evaluating a quality measurement system. These expectations bear directly on the definition of quality. Initially, the Institute of Medicine (IOM) defined quality as the degree to which “health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge.”³³ Subsequently, the IOM published *Crossing the Quality Chasm* and presented what may be construed as a more comprehensive framework of health care quality.³⁴ Pillars of this framework extend beyond evidence-based medicine to include care based on continuous healing relationships, shared decision making, and systems that are safe and not wasteful. In addition, the IOM advocates care that is equitable and not biased by ethnicity, gender, or socioeconomic position.³⁵

Isolated steps are being taken to extend quality measurement in the direction suggested by the IOM. For example, NCQA accreditation relies, in part, on the Consumer Assessment of Health Plans (CAHPS®) survey.^{36,37} While “balanced scorecards” incorporate the consumer’s perspective, many were developed for financial or marketing reasons and they often, if not usually, fail to capture the spirit of the IOM reports.³⁸ Currently, no measurement system encompasses the entire quality spectrum. These conceptual and technical problems suggest that pursuit of realistic excellence is more appropriate than striving for artificial perfection as reflected by an imperfect measurement system.³⁹

Overall, we need a paradigm shift in quality measurement, not just an incremental advance.⁴⁰ Change of this magnitude demands more than improved methodological approaches to measure development, solutions to recognized technical problems, and inclusion of more conditions.

CAHPS® is a registered trademark of the Agency for Healthcare Research and Quality (AHRQ).

Conclusion: Four Principles

Neither the published medical literature nor material from quality rating organizations directly addressed the Special Study's focus on hospitals that are high performing in their management of the three conditions from the 7th SoW. However, with the concurrence of the review findings, the hospital meetings, and conversations with consultants and among the collaborators to this Special Study, it has become possible to identify four principles that will guide HSAG in proceeding to a final definition of high performer and the development of the algorithm to identify them. These principles are:

1. Hospitals that provide the same service should be held to the same standard, regardless of size or geographic location.
2. Performing well on an indicator on which others perform poorly (a "hard" indicator), while performing as well as any on an indicator on which others do well (an "easy" indicator), is basic to a definition of high performance.
3. Performing well on "hard" indicators is more important than performing well on "easy" indicators.
4. High performance must incorporate sustained excellence over time.

The information synthesized from the current review has reaffirmed CMS' position that there is an underlying need within the health care system to design a scientifically sound and credible structure by which high performing health care providers can be defined, identified, and studied. Furthermore, the information synthesized from the current review has established a high degree of feasibility to design such a structure. The goal of this special study is to provide direction and opportunity to move the health care field one step closer to understanding the mechanisms by which organizational characteristics mediate quality performance. Such an understanding would afford additional opportunity for QIOs to provide assistance for improving health care quality to all stakeholders by replicating successful models.

This section presents the following tables:

Table 3-1—Literature Describing Methods Used to Identify High Performers, by Health Care Setting.....	3-2
Table 3-2—Literature Evaluating Various Statistical Profiling Methods.....	3-6
Table 3-3—7th SoW Clinical Conditions Used in Measuring Institutional or Practitioner Performance in Hospitals.....	3-8
Table 3-4—7th SoW Clinical Conditions Used in Measuring Institutional or Practitioner Performance, by Other Settings.....	3-9
Table 3-5—Quality Improvement Dimensions Affecting Quality Performance Addressed in Literature and Organization Material.....	3-10

Table 3-1. Literature Describing Methods Used to Identify High Performers, by Health Care Setting

Reference (Pub. Date)	Sample/ Setting	Study Purpose	Methods	Results (Relevance to High Performer Algorithm Development) ¹
Hospitals				
Ansari (1999) ⁵⁰	Patents with prostatectomy at 21 non-teaching and 15 teaching hospitals/ Australia	Compare crude and adjusted in-hospital mortality rates after prostatectomy between teaching and non-teaching hospitals, using hospital discharge data.	Mortality rates from 15 teaching hospitals were used as the comparison group in multivariate analysis. Comparisons among the non-teaching hospitals were based on crude and adjusted odds ratios (OR) and 95% confidence intervals (CI) derived using univariate and multivariate logistic regression.	The overall crude mortality rates among the 21 hospitals achieved borderline significance; these differences were no longer significant after adjustment for patient and hospital characteristics. On crude analysis of mortality rates, four hospitals were initially identified as 'low' outlier hospitals; after adjustment, none of these remained outside the 95% CI, whereas a new hospital emerged as a 'high' outlier.
Austin (2002) ⁴¹	152 acute care hospitals with 18,850 discharges 1997–98/ Ontario, Canada	Compare 4 different Bayesian methods for classifying hospitals as outcomes outliers, using 30-day hospital-level mortality rates for a cohort of acute myocardial infarction patients as a test case.	Bayesian random intercept model used to: (1) determine the probability that a hospital's mortality rate for an average patient exceeded a specified threshold; (2) rank hospitals according to their mortality rate for an average patient; (3) determine the probability that a hospital's standardized mortality ratio (smr) exceeded a specified threshold; (4) rank hospitals according to their smr. Each hospital was classified as an outlier, low outlier, or neither for each of the 4 methods. Kappa statistic was used to test agreement between methods.	In most of the scenarios examined, there was only marginal agreement between the different methods. In only 4 of 19 comparisons was there good agreement between the different methods ($0.40 < \text{or} = \text{kappa} < \text{or} = 0.75$). Methods based on ranking institutions based on mortality for an average patient were relatively insensitive to differences between hospitals; ranking based on smr allows classification of several hospitals as performance outliers.
Bronskill (2002a) ⁴⁵	32 hospitals CABG patients/New York State and 156 VA hospitals/ nationwide	Examine and classify provider performance over time using examples from cardiology and mental health.	Longitudinal profiling with 1991–1995 data on CABG and 1993–1998 data on mental health patients, using hierarchical generalized linear models (HGLM) to model variations among cardiac providers and mental health networks.	Able to profile on physicians having a small as well as a large number of annual cases, and profile mental health networks on mental health readmissions. Able to generate information on the provider-specific evolution of performance and compare to peers.
Bronskill (2002b) ⁵⁷	66,399 patients from non-Federal acute care hospitals/ 7 states	Examine the extent to which factors other than patient characteristics contribute to post-acute service use (PASU) within 30 days following hospital discharge for AMI.	Applied a published risk-adjustment algorithm, based on patient characteristics and process of care measures, to merged datasets from the 1994–95 Cooperative Cardiovascular Project (CCP), Medicare service use, and hospital characteristics.	Model enabled comparisons of hospital PASU rates while controlling for hospital differences in the type and number of patients treated, and allowing for random between-hospital variation. For-profit ownership and provision of home health services by the discharging hospital were significant predictors of PASU.

Table 3-1. Literature Describing Methods Used to Identify High Performers, by Health Care Setting

Reference (Pub. Date)	Sample/ Setting	Study Purpose	Methods	Results (Relevance to High Performer Algorithm Development) ¹
Chen (1999) ⁵¹	4,672 hospitals/ nationwide	To compare <i>US News & World Report</i> top-rated hospitals with all other hospitals in the Cooperative Cardiovascular project.	Using multivariate techniques, Medicare beneficiaries with AMI (1994–94) were compared for 3 hospital types: top-ranked hospitals; similarly equipped hospitals; and the non-similarly equipped hospitals. Outcomes included 30-day mortality; the use rates of aspirin, beta-blockers, and reperfusion; and the relation of differences in rates of therapy to short-term mortality.	Admission to a hospital ranked high on the list of “America’s Best Hospitals” was associated with lower 30-day mortality. The survival advantage associated with admission to top-ranked hospitals was less strong after adjustment for factors including the use of aspirin and beta-blockers, suggesting a substantial portion of the survival advantage may be associated with these hospitals’ higher rates of use of aspirin and beta-blocker therapy.
Griffith (2002a) ⁴⁹	Datasets from about 3,000 hospitals/ nationwide	Evaluate the content validity, reliability and sensitivity, and the rank order comparisons of a set of 9 performance measures derived from Medicare reports.	Nine derived utilization measures from Medicare reports 1996–1998 (data set prepared by Solucient). Content validity assessed by definition review and relationship to generally accepted concepts. Reliability, sensitivity, and stability assessed with descriptive techniques, ranking, examination of year-to-year correlation, and construction of 2 and 3 year moving averages.	The most useful measures for comparing hospitals were related to asset turnover, cash flow, mortality, complications, and cost per case. These measures were found to be independent of common hospital characteristics and county demographic characteristics, and could be used to rank hospitals by quintile against the national distribution.
Griffith (2002b) ²⁴	742 matched hospitals/ nationwide	Compare a set of 7 non-federal general hospital performance measures derived from Medicare against JCAHO accreditation measures.	Using multivariate and correlation analyses, data sets from Medicare and JCAHO 1996–1998 used to compare 7 measures selected from Medicare reports by Solucient and JCAHO measures of decisions and overall evaluation scores (OES) made publicly available.	Only 2 of Solucient measures had a significant association with OES scores and 1 measure was in an unexpected direction. Of 315 Pearson correlations, only 9 explained more than 1 percent of the variance. Only 1 significant difference (in contrary to assumed direction) was found between extreme quartiles and octiles of Solucient versus JCAHO scores.
Jacobs (2001) ⁴³	Hospital cost dataset/UK	Compare hospital efficiency rankings from 3 cost indices using data envelopment analysis (DEA) and stochastic frontier analysis (SFA).	Using 1996 data from the National Health Service, applied DEA (a linear programming method) and SFA (econometric regression technique) to test 5 specified models of efficiency rankings, using the same variable set from 3 cost indices developed by the Department of Health.	There was a high degree of correlation between the cost indices and 5 SFA specifications, and within DEA and SFA models. Methods each have particular strengths and weaknesses and potentially measure different aspects of efficiency. Differences in efficiency scores across different methods may be due to random “noise” and reflect data deficiencies. Longitudinal examination of efficiency would allow better assessment of random “noise.”
Jencks (2003) ⁵⁶	Hospital in- patients/ nationwide	Track national and state-level changes in performance on 22 quality indicators of care for Medicare beneficiaries.	Using 1998–1999 (baseline) and 2000–2001 (follow-up) data, descriptive techniques used to compare within and between state performances, and to rank states based on overall performance on all 22 indicators.	The median state’s performance improved from baseline to follow-up on 20 of the 22 indicators. For all but one indicator, absolute improvement was greater in states in which performance was low at baseline than those in which it was high at baseline. When states were ranked on each indicator, the state’s average rank was highly stable over time.

Table 3-1. Literature Describing Methods Used to Identify High Performers, by Health Care Setting

Reference (Pub. Date)	Sample/ Setting	Study Purpose	Methods	Results (Relevance to High Performer Algorithm Development) ¹
Jiang (2001) ⁵²	Over 500 hospitals/ New York State	Describe how the Healthcare Association of New York produced the Comparative Hospital Report for members, Healthcare Cost and Utilization Project (HCUP) quality indicators in 3 areas (patient outcomes, utilization of selected procedures, access to primary care) developed by AHRQ.	33 measures applied to 1997 N.Y. State hospital inpatient discharge abstract data. A risk adjustment model was constructed for each complication measure, and indirect standardization was used for utilization and access to care to adjust for differences in patient demographics and payer status.	Results were presented to hospitals in graphic format, with comparisons to peers, region, and state. Hospitals found the report useful for internal quality assessment and improvement, as a way to identify community health care needs. Indicators limited by lack of valid measures and low incidence rates for some outcomes.
Kiefe (1998) ³⁹	106 hospitals/ Alabama	Demonstrate application of ABC methods to process of care measures.	Using pared-mean methods to determine “top performance” defined as the mean of the best care achieved for at least 10% of the population.	The limitation of low denominators was addressed using Bayesian estimator technique, yielding a Bayesian estimator adjusted performance statistic. To identify providers who perform at a high level on several indicators for an aggregate ABC, a formula for calculating minimum sufficient denominator (MSD) was derived for application to each indicator in a given dataset.
Morrissey (1998) ⁵³	3,258 hospitals/ nationwide	Rank top 100 high achieving hospitals by a health care information company, HCIA and Mercer.	Used 1997 Medicare cost and discharge data to measure 8 indicators (3 fiscal, 3 operational, and 2 clinical) against past 2 years benchmarks. Analytic methods not described.	Data not available. Fiscal and operational indicators stabilized with opportunity to improve based on improving clinical processes with use of clinical guidelines.
Morrissey (1999) ⁵⁴	2,946 acute care hospitals/ nationwide	Rank top 100 high achieving hospitals by a health care information company, HCIA.	Used 1998 Medicare cost and discharge data to measure 8 indicators (3 fiscal, 3 operational, and 2 clinical) against past 2 years benchmarks. Analytic methods not described.	Data not available. Benchmarks, which had improved for 5 years, did not improve from the previous year. The gap among the top 100 hospitals narrowed. Hospitals found to be less profitable, with higher occupancy and higher complications; but use of clinical protocols and other management systems significantly reduced complications.
O’Neal (2002) ⁴⁴	62 university teaching hospitals/ national database	Rate the technical efficiency of mechanical ventilation in University Hospital Consortium hospitals.	Used VRS DEA modeling as a benchmarking technique to assess technical efficiency of hospitals. The analyses used 4 resource-related input variables and 3 patient outcome variables to generate 2 models.	10 hospitals were considered efficient and 52 were considered inefficient. All inputs demonstrated excessive resource utilization among inefficient hospitals.
Richman (1999) ⁴⁷	239 Maternity Units/Nova Scotia	Determine an achievable benchmark cesarean section (c/s) rate.	C/S rate benchmarks were established for 3 size categories of maternity units by successively setting the 75th, 50th, and 25th percentiles within each unit. Units with higher c/s rates than the assigned statistical rate limit were set to the assigned rate limit; those with lower c/s rates were not changed.	50th percentile c/s rate medians were similar for 3 groupings. The interquartile differences increased when there were more units within a category.

Nursing Homes				
Berlowitz (2002) ⁴²	108 nursing homes/ national for-profit nursing home chain	Examine the use of Bayesian hierarchical modeling in profiling nursing homes on their rate of pressure ulcer development.	Using Minimum Data Set data (1997–1998), nursing homes were compared on their performance on risk-adjusted rates of pressure ulcer development calculated using standard statistical techniques and Bayesian hierarchical modeling.	The range of risk-adjusted rates among nursing homes was 0% to 14.3% using standard methods, and 1.0% to 4.8% using Bayesian analysis. Fifteen nursing homes were designated as outliers based on their z scores, and two were outliers using Bayesian modeling. Only one nursing home had greater than a 50% probability of having a true rate of ulcer development exceeding 4%.
Physicians				
MacLean (2000) ⁴⁶	1,355 rheumatoid arthritis patients/ nationwide	Assess the quality of the health care that rheumatoid arthritis patients receive for their arthritis, comorbid diseases, and health care maintenance and to determine the effect of patterns of specialty care on quality.	Using administrative data from a U.S. insurance company 1991–1995, developed process measures that had links to evidence-based outcomes, in 3 dimensions (arthritis, comorbid disease, health maintenance). Calculations were made for individual, within domain, and aggregate measures. Comparative analysis was by pattern of specialty care.	Health care quality was suboptimal for arthritis, comorbid disease, and health care maintenance. Patterns of care that included relevant specialists were associated with substantially higher quality across all dimensions. Patterns that included generalists were associated with substantially higher quality health care maintenance than patterns that included neither a generalist nor a relevant specialist.
Parkerton (2003) ⁵⁵	194 HMO family practitioners and general internists/ Calif.	To determine the consistency and reliability of 4 HEDIS measures of primary care physician performance measures: cancer screening, diabetes management, patient satisfaction, and ambulatory costs.	Administrative data on physician practice and performance were assessed with multiple regression, controlling for physician and physician panel characteristics, and analysis of variance.	Each performance measure was significantly related to 1 or 2 of the other measures. Physicians excelling on one measure did not perform better than average on all. Physician performance ranks did not cluster. Predictive value low for performance levels, given measures on some of the indicators. Aggregate measures of different practice aspects reduce performance to a mean rather than identifying effective or ineffective practice. Validity of measures to assess total practice of individual physicians is uncertain.
Wahls (2000) ⁴⁸	Marshfield Clinic/Wisc.	Identify the clinic’s “better performing” physicians, or “master clinicians,” in terms of productivity, professional satisfaction, and patient satisfaction. Identify the “best practices” used by these better performers. Develop a training curriculum to disseminate the best practice information.	Developed internal benchmarks followed by interviews of physician sample to discover best practices of high performers. Best practices organized into a set of clinical competencies, and training curriculum developed and delivered. Evaluation planned by tracking productivity and patient satisfaction of trained physicians vs. comparison group.	Developed a reproducible internal benchmarking method to identify best performers of medical groups, using patient satisfaction survey and productivity data.

Table 3-2. Literature Evaluating Various Statistical Profiling Methods

Reference (Pub. Date)	Purpose	Argument/Critique
Allison (2002) ¹¹	Discuss the appropriate application of inferential statistics to practice profiles and other measures of care.	Review of benchmarking methods, use of large cross-sectional datasets, and essential attributes for a quality profile. Discussion of statistical considerations with clustered data, aggregate quality measures, and generalized linear models and hierarchical modeling for clustered data. Authors assert that the use of inferential statistics in profiling requires further study.
Christiansen (1997) ⁶²	Review and compare existing statistical methods for profiling health care providers.	Standard statistical procedures ignore units with small caseloads. Hierarchical models require no minimum size, provided the ensemble of all providers analyzed has adequate data. The use of ensemble data allows each provider's performance to be calculated yielding profiles meaningful for health care policy.
Green (1997) ⁶³	Critique methods used to produce an "index of hospital quality" for rating performance of "America's Best Hospitals," by US News & World Report (1995).	Review of the 1995 methods found that hospital coding of medical specialties was ambiguous and limited; use of total bed count by specialty and identification of technology services was biased; identification of teaching status was incomplete; measurement of procedure volume as well as reliance on using only Medicare data to identify procedures was inadequate; mortality data was unreliable; and physician ranking of best hospitals was biased. Measurement systems need to consider the processes of care.
Lee (2002) ⁶⁴	Demonstrate use of JCAHO ORYX data to compare and evaluate organizational performance.	Tutorial on the construction, use, and interpretation of comparison charts, using ORYX cross-sectional data and descriptive techniques, to compare hospital performance.
Mant (1996) ⁵⁹	Explore the limitations of using outcomes measures to compare provider performance.	Valid process of care measures are more sensitive than measures of outcome in detecting differences in quality. However, outcomes may be useful if the potential effect is large; power of the study is adequate; and selection bias is taken into account, preferably with randomized controlled studies. Monitoring outcomes of health care quality is useful when how you do something is as important as what you do, when process measures are invalid or impractical, and when the overall effectiveness of an intervention is critically

Table 3-2. Literature Evaluating Various Statistical Profiling Methods

Reference (Pub. Date)	Purpose	Argument/Critique
		dependent upon its complication rate.
Normand (2002) ⁶¹	Review approaches to sample size determination to compare providers when designing observational health care quality studies.	There are several approaches available to researchers to analyze dependent binary data. Marginal and hierarchical modeling is examined. If researchers are interested in comparing clusters, hierarchical models are preferred for determining sample size.
Powell (2003) ⁵⁸	Identify strengths and weaknesses of using comparative data to assess quality of health care.	Four main issues affect the interpretation of comparative routine outcomes data: measurement properties, controlling for case mix and other relevant factors, coping with chance variability, and data quality.
Shahian (2001) ⁶⁰	Identify flaws of statistical methods used to produce surgery report cards.	The statistical methodology underlying current report cards is flawed, and does not justify the degree of accuracy presented to the public. Existing risk-adjustment methods are imprecise, and compounded when the results of patient-level models are aggregated and used inappropriately to assess provider performance. Authors advocate the use of hierarchical or multilevel statistical models to address these problems. They also advocate using report formats that emphasize the statistical uncertainty of the results.

Table 3-3. 7th SoW Clinical Conditions Used in Measuring Institutional or Practitioner Performance in Hospitals		
Hospital Conditions and Measures	Number of Studies*	Reference # of Studies Addressing SoW Indicators
Myocardial Infarction	11	
Early administration of aspirin	5	56, 66, 67, 70, 71
Aspirin at discharge	1	56
Early administration of beta blocker	5	56, 66, 67, 70, 71
Beta blocker at discharge	1	56
ACE inhibitor at discharge for systolic dysfunction	4	56, 66, 67, 70,
Time to initiation of reperfusion therapy	5	56, 66, 67, 70, 71
Smoking cessation counseling	2	56, 70
Heart Failure	7	
Evaluation of left ventricular function	1	56
ACE inhibitor at discharge for systolic dysfunction	2	56, 72
Discharge instructions	0	
Smoking cessation counseling	0	
Pneumonia	7	
Blood culture before antibiotics	3	56, 76, 79
Time to initial antibiotic administration	5	56, 76, 77, 78, 79
Administration of antibiotics consistent with guidelines	1	56
Pneumococcal vaccination, inpatient	2	56, 81
Influenza immunization, inpatient	1	56
Oxygenation assessment within 24 hours of admission	0	0
Smoking cessation counseling	0	0
Surgical Wound Infection	0	0
Correct prophylactic antibiotic	0	0
Correct timing of antibiotic administration	0	0
Correct duration of antibiotic administration	0	0
* A study may have addressed more than 1 condition.		

Table 3-4. 7th SoW Clinical Conditions Used in Measuring Institutional or Practitioner Performance, by Other Settings

Setting, Conditions, and Measures	Number of Studies*	Reference # of Studies Addressing SoW Indicators
Nursing Homes[†]	10	
Pressure ulcer	6	82, 83, 84, 85, 86, 118
Activities of daily living	0	
Confusion	1	90
Infection control	0	
Pain	0	
Physical restraints	2	88, 89
Walking	0	
Home Care	1	
Activities of daily living	0	
Urinary tract infections	0	
Dyspnea	0	
Patient re-hospitalization	1	91
Rehabilitation of speech and language	0	
Anxiety	0	
Behavioral symptoms	0	
Physicians	4	
Diabetes screening – retinopathy	2	92, 94
Hemoglobin A, glycosylated	2	92, 93
Lipid profile	2	92, 94
Mammography	1	95
Pneumococcal immunization	0	
* A study may have addressed more than 1 condition.		

[†] These are the original quality measures used in the national quality initiative. An enhanced set of measures went into effect January 22, 2004.

Table 3-5. Quality Improvement Dimensions Affecting Quality Performance Addressed in Literature and Organization Material

Dimensions	Shortell SM (1985) ⁹⁹	Shortell SM, O'Brien JL, et al. (1995) ¹⁰⁰	Shortell SM, Levin DZ, et al. (1995) ¹⁰¹	Shortell SM, Bennett CL, Byck GR (1998) ¹⁰²	Berwick DM; James B, Coyle, MJ (2003) ¹⁰³	Scott J, Schenkman M et al. (2003) ¹⁰⁴	Aiken LH, Sochalski J (1997) ¹⁰⁵	Scott J, Vojir, C (2003) ¹¹⁷	Levesque DA, Prochaska JM, et al. (2001) ¹⁰⁶	JCAHO ¹⁰⁷	ISO for Health Care Organizations ⁹⁸	Malcolm Baldrige Award ⁹⁷
Strategic			X		X					X	X	X
Structural		X	X						X	X	X	
Cultural	X	X	X	X			X	X		X	X	X
Leadership	X		X	X	X	X				X	X	X
Systems										X	X	X
Customer and Market								X	X	X	X	X
Measurement, Analysis, and Knowledge Management				X	X					X	X	X
Human Resources			X			X		X		X	X	X
Business Results										X	X	X

4. References and Endnotes

1. Braun BL, Kind EA, Fowles JB, Suarez WG. Consumer response to a report card comparing healthcare systems. *Am J Manag Care*. 2002;8(6):522–528.
2. Lansky D. Improving quality through public disclosure of performance information. *Health Affairs*. 2002;21(4):52–62.
3. McGlynn EA, Kerr EA, Adams J, Keesey J, Asch SM. Quality of health care for women: A demonstration of the QA Tools system. *Med Care*. 2003;41(5):616–25.
4. Puig-Junoy J. Managing risk selection incentives in health sector reforms. *Int J Health Plann Manage*. 1999;14(4):287–311.
5. Casalino LP. The unintended consequences of measuring quality on the quality of medical care. *N Engl J Med*. 1999;341(15):1147–1150.
6. Hofer TP, Hayward RA, Greenfield S, Wagner EH, Kaplan SH, Manning WG. The unreliability of individual physician "report cards" for assessing the costs and quality of care of a chronic disease. *JAMA*. 1999;281(22):2098–2105.
7. Allison JJ, Hardin M. Electronic medical record databases and quality improvement: Concepts and technical issues. In: Carter JH, ed. *Electronic Medical Records: A Guide for Clinicians and Administrators*: American College of Physicians Press—American Society of Internal Medicine; 2001.
8. Desirable Attributes of Performance Measures: A Consensus Document. The American Medical Association, The Joint Commission on Accreditation of Healthcare Organizations, and The National Committee for Quality Assurance; April 19, 1999.
9. Normand ST, Glickman ME, Gastonis CA. Statistical methods for profiling providers of medical care: Issues and applications. *J Am Stat Assoc*. 1997;92(439):803–814.
10. Landrum MB, Bronskill SE, Normand ST. Analytic methods for constructing cross-sectional profiles of health care providers. *Health Serv Outcomes Res Method*. 2000;1(1):23–47.
11. Allison JJ, Calhoun JW, Wall TC, et al. Optimal reporting of health care process measures: Inferential statistics as help or hindrance? *Manag Care Q*. Autumn 2000;8(4):1–10.
12. Anderson J, Carling K, Mattson S. Random ranking of hospitals is unsound. *Chance*. 1998;11(3):34–37, 39.
13. Zaslavsky AM. Statistical issues in reporting quality data: Small samples and casemix variation. *Int J Qual Health Care*. 2001;13(6):481–488.
14. Donner A, Klar N. Methods for comparing event rates in intervention studies when the unit of allocation is a cluster. *Am J Epidemiology*. 1994;140(3):279–289; discussion 300–301.
15. Greenfield S, Kaplan SH, Kahn R, Ninomiya J, Griffith JL. Profiling care provided by different groups of physicians: Effects of patient case-mix (bias) and physician-level clustering on quality assessment results. *Ann Intern Med*. 2002;136(2):111–121.
16. Bravo G, Potvin L. Estimating the reliability of continuous measures with Cronbach's alpha or the intraclass correlation coefficient: Toward the integration of two traditions. *J Clin Epidemiology*. 1991;44(4–5):381–390.

17. 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):793–796.
18. Katz JN. Patient preferences and health disparities. *JAMA*. 2001;286(12):1506–1509.
19. Naylor CD. Public profiling of clinical performance. *JAMA*. 2002;287(10):1323–1325.
20. Chassin MR. Achieving and sustaining improved quality: Lessons from New York State and cardiac surgery. *Health Affairs*. 2002;21(4):40–51.
21. Gandhi TK, Francis EC, Puopolo AL, Burstin HR, Haas JS, Brennan TA. Inconsistent report cards: Assessing the comparability of various measures of the quality of ambulatory care. *Med Care*. 2002;40(2):155–165.
22. Donabedian A. The Definition of Quality and Approaches to Its Assessment. *Explorations in Quality Assessment and Monitoring*. Vol 1. Ann Arbor, MI: Health Administration Press; 1980.
23. Dean Beaulieu N, Epstein AM. National Committee on Quality Assurance health-plan accreditation: Predictors, correlates of performance, and market impact. *Med Care*. 2002;40(4):325–337.
24. Griffith JR, Knutzen SR, Alexander JA. Structural versus outcomes measures in hospitals: A comparison of Joint Commission and Medicare outcomes scores in hospitals. *Qual Manag Health Care*. 2002;10(2):29–38.
25. Needleman J, Buerhaus P, Mattke S, Stewart M, Zelevinsky K. Nurse-staffing levels and the quality of care in hospitals. *N Engl J Med*. 2002;346(22):1715–1722.
26. Meyer GS, Massagli MP. The forgotten component of the quality triad: Can we still learn something from "structure"? *Jt Comm J Qual Improv*. 2001;27(9):484–493.
27. Clancy CM, Eisenberg JM. Outcomes research: Measuring the end results of health care. *Science*. 1998;282(5387):245–246.
28. Iezzoni LI. The risks of risk adjustment. *JAMA*. 1997;278(19):1600–1607.
29. Nash DB. Quality of measurement or quality of medicine? *JAMA*. 1995;273(19):1537–1538.
30. 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):1509–1514.
31. Shekelle PG. Are appropriateness criteria ready for use in clinical practice? *N Engl J Med*. 2001;344(9):677–678.
32. Shekelle PG, Kahan JP, Bernstein SJ, Leape LL, Kamberg CJ, Park RE. The reproducibility of a method to identify the overuse and underuse of medical procedures. *N Engl J Med*. 1998;338(26):1888–1895.
33. Institute of Medicine. *Medicare: A Strategy for Quality Assurance, Vol. One*. Washington, D.C.: National Academy Press; 1990.
34. Institute of Medicine. *Crossing the Quality Chasm*. Washington, D.C.: National Academy Press; 2001.
35. Institute of Medicine. *Guidance for the National Healthcare Disparities Report*. Washington, D.C.: National Academy Press; 2002.

36. *The National Committee for Quality Assurance. The State of Managed Care Quality 2001.* Available at <http://www.healthchoices.org/about/SOMCReport2001.pdf>.
37. Hibbard JH, Berkman N, McCormack LA, Jael E. The impact of a CAHPS report on employee knowledge, beliefs, and decisions. *Med Care Res Rev.* 2002;59(1):104–116.
38. Oliveira J. The balanced scorecard: An integrative approach to performance evaluation. *Healthc Financ Manage.* 2001;55(5):42–46.
39. Kiefe CI, Weissman NW, Allison JJ, Farmer R, Weaver M, Williams OD. Identifying achievable benchmarks of care: Concepts and methodology. *Int J Qual Health Care.* 1998;10(5):443–447.
40. Kuhn TS. *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press; 1996.
41. Austin PC. A comparison of Bayesian methods for profiling hospital performance. *Med Decis Making.* 2002;22(2):163–172.
42. Berlowitz DR, Christiansen CL, Brandeis GH, et al. Profiling nursing homes using Bayesian hierarchical modeling. *J Am Geriatr Soc.* 2002;50(6):1126–1130.
43. Jacobs R. Alternative methods to examine hospital efficiency: Data envelopment analysis and stochastic frontier analysis. *Health Care Manag Sci.* 2001;4(2):103–115.
44. O'Neal PV, Ozcan YA, Ma Y. Benchmarking mechanical ventilation services in teaching hospitals. *J Med Syst.* 2002;26(3):227–240.
45. Bronskill SE, Normand SL, Landrum MB, Rosenheck RA. Longitudinal profiles of health care providers. *Stat Med.* 2002;21(8):1067–1088.
46. MacLean CH, Louie R, Leake B, et al. Quality of care for patients with rheumatoid arthritis. *JAMA.* 2000;284(8):984–992.
47. Richman VV. Setting goals for reductions in Canadian cesarean delivery rates: Benchmarking medical practice patterns. *Am J Obstet Gynecol.* 1999;181(3):635–637.
48. Wahls TL, Futz DN. The master clinician project. *J Ambul Care Manag.* 2000;23(4):9–21.
49. Griffith JR, Alexander JA, Jelinek RC. Measuring comparative hospital performance. *J Healthc Manag.* 2002;47(1):41–57.
50. Ansari MZ, Ackland MJ, Jolley DJ, Carson N, McDonald IG. Inter-hospital comparison of mortality rates. *Int J Qual Health Care.* 1999;11(1):29–35.
51. Chen J, Radford MJ, Wang Y, Marciniak TA, Krumholz HM. Do "America's Best Hospitals" perform better for acute myocardial infarction? *N Engl J Med.* 1999;340(4):286–292.
52. Jiang HJ, Ciccone K, Urlaub CJ, Boyd D, Meeks G, Horton L. Adapting the HCUP QIs for hospital use: The experience in New York State. *Jt Comm J Qual Improv.* 2001;27(4):200–215.
53. Morrissey J. All benchmarked out: Even the top 100 hospitals can't find many more ways to be more productive. *Mod Healthc.* 1998;28(49):38–40, 42–34, 46.
54. Morrissey J. Top 100 hospitals. *Mod Healthc.* 1999;29(50):20–24, 28–29.
55. Parkerton PH, Smith DG, Belin TR, Feldbau GA. Physician performance assessment: nonequivalence of primary care measures. *Med Care.* 2003;41(9):1034–1047.
56. Jencks SF, Huff ED, Cuedon T. Change in the quality of care delivered to Medicare beneficiaries, 1998–1999 to 2000–2001. *JAMA.* 2003;289(3):305–312.

57. Bronskill SE, Normand SL, McNeil BJ. Post-acute service use following acute myocardial infarction in the elderly. *Health Care Financ Rev.* 2002;24(2):77–93.
58. Powell AE, Davies HTO, Thomson RG. Using routine comparative data to assess the quality of health care: Understanding and avoiding common pitfalls. *Qual Saf Health Care.* 2003; 12(2):122–128.
59. Mant J, Hicks NR. Assessing quality of care: What are the implications of the potential lack of sensitivity of outcome measures to differences in quality? *J Eval Clin Pract.* 1996;2(4):243–248.
60. Shahian DM, Normand SL, Torchiana DF, et al. Cardiac surgery report cards: Comprehensive review and statistical critique. *Ann Thorac Surg.* 2001;72(6):2155–2168.
61. Normand SL, Zou KH. Sample size considerations in observational health care quality studies. *Stat Med.* 2002;21(3):331–345.
62. Christiansen CL, Morris CN. Improving the statistical approach to health care provider profiling. *Ann Intern Med.* 1997;127(8 Pt 2):764–768.
63. Green J, Wintfeld N, Krasner M, Wells C. In search of America's best hospitals: The promise and reality of quality assessment. *JAMA.* 1997;277(14):1152–1155.
64. Lee KY, McGreevey C. Using comparison charts to assess performance measurement data. *Jt Comm J Qual Improv.* 2002;28(3):129–138.
65. Rosenthal GE, Baker DW, Norris DG, Way LE, Harper DL, Snow RJ. Relationships between in-hospital and 30-day standardized hospital mortality: Implications for profiling hospitals. *Health Serv Res.* 2000;34(7):1449–1468.
66. Krumholz HM, Rathore SS, Chen J, Wang Y, Radford MJ. Evaluation of a consumer-oriented Internet health care report card: The risk of quality ratings based on mortality data. *JAMA.* 2002;287(10):1277–1287.
67. 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):1256–1262.
68. Philbin EF, Rocco TA, Lindenmuth NW, Ulrich K, McCall M, Jenkins PL. The results of a randomized trial of a quality improvement intervention in the care of patients with heart failure. The MISCHF Study Investigators. *Am J Med.* 2000;109(6):443–449.
69. Baker DW, Einstadter D, Thomas CL, Husak SS, Gordon NH, Cebul RD. Mortality trends during a program that publicly reported hospital performance. *Med Care.* 2002;40(10):879–890.
70. Mehta RH, Montoye CK, Gallogly M, et al. Improving quality of care for acute myocardial infarction: The Guidelines Applied in Practice (GAP) Initiative. *JAMA.* 2002; 287(10):1269–1276.
71. Willison DJ, Soumerai SB, Palmer RH. Association of physician and hospital volume with use of aspirin and reperfusion therapy in acute myocardial infarction. *Med Care.* 2000;38(11):1092–1102.
72. Weil E, Tu JV. Quality of congestive heart failure treatment at a Canadian teaching hospital. *CMAJ.* 2001;165(3):284–287.
73. Shen YC. The effect of hospital ownership choice on patient outcomes after treatment for acute myocardial infarction. *J Health Econ.* 2002;21(5):901–922.

74. Welke KF, BootsMiller BJ, McCoy KD, et al. What factors influence provider knowledge of a congestive heart failure guideline in a national health care system? *Amer J Med Qual.* 2003;18(3):122–127.
75. Holmboe ES, Meehan TP, Radford MJ, Wang Y, Krumholz HM. What's happening in quality improvement at the local hospital: A state-wide study from the Cooperative Cardiovascular Project. *Amer J Med Qual.* 2000;15(3):106–113.
76. Chu LA, Bratzler DW, Lewis RJ, et al. Improving the quality of care for patients with pneumonia in very small hospitals. *Arch Intern Med.* 2003;163(3):326–332.
77. Lawrence SJ, Shadel BN, Leet TL, Hall JB, Mundy LM. An intervention to improve antibiotic delivery and sputum procurement in patients hospitalized with community-acquired pneumonia. *Chest.* 2002;122(3):913–919.
78. Meehan TP, Weingarten SR, Holmboe ES, et al. A statewide initiative to improve the care of hospitalized pneumonia patients: The Connecticut Pneumonia Pathway Project. *Amer J Med.* 2001;111(3):203–210.
79. Fine JM, Fine MJ, Galusha D, Petrillo M, Meehan TP. Patient and hospital characteristics associated with recommended processes of care for elderly patients hospitalized with pneumonia: Results from the Medicare quality indicator system pneumonia module. *Arch Intern Med.* 2002;162(7):827–833.
80. Arozullah AM, Henderson WG, Khuri SF, Daley J. Postoperative mortality and pulmonary complication rankings: How well do they correlate at the hospital level? *Med Care.* 2003;41(8):979–991.
81. Stephens MK, Cochran RF, Schade CP. Improving medical care for West Virginia seniors. *W V Med J.* 2001;97(4):188–193.
82. Berlowitz DR, Brandeis GH, Anderson JJ, et al. Evaluation of a risk-adjustment model for pressure ulcer development using the Minimum Data Set. *J Am Geriatr Soc.* 2001;49(7):872–876.
83. Berlowitz DR, Bezerra HQ, Brandeis GH, Kader B, Anderson JJ. Are we improving the quality of nursing home care: The case of pressure ulcers. *J Am Geriatr Soc.* 2000;48(1):59–62.
84. Bates-Jensen BM, Cadogan M, Jorge J, Schnelle JF. Standardized quality-assessment system to evaluate pressure ulcer care in the nursing home. *J Am Geriatr Soc.* 2003;51(9):1194–1202.
85. Bates-Jensen BM, Cadogan M, Osterweil D, et al. The minimum data set pressure ulcer indicator: Does it reflect differences in care processes related to pressure ulcer prevention and treatment in nursing homes? *J Am Geriatr Soc.* 2003;51(9):1203–1212.
86. Berlowitz DR, Young GJ, Hickey EC, et al. Quality improvement implementation in the nursing home. *Health Serv Res.* 2003;38(1 Pt 1):65–83.
87. Courtney BM, Spencer L. What's best? Clinical indicators of quality in residential aged care facilities. *Collegian.* 2000;7(2):14–19.
88. Castle NG. Nursing homes with persistent deficiency citations for physical restraint use. *Med Care.* 2002;40(10):868–878.

89. Carroll-Solomon PA, Christian V, Denny DS, Nordan VN, Therriault MF, Van Wicklen R. Preserving residents' rights in long-term care settings: A values-based approach to restraint reduction. *J Healthc Qual.* 2000;22(4):10–19.
90. Sloane PD, Mitchell CM, Weisman G, et al. The Therapeutic Environment Screening Survey for Nursing Homes (TESS-NH): An observational instrument for assessing the physical environment of institutional settings for persons with dementia. *J Gerontol B Psychol Sci Soc Sci.* 2002;57(2):S69–78.
91. Rosati RJ, Huang L, Navaie-Waliser M, Feldman PH. Risk factors for repeated hospitalizations among home healthcare recipients. *J Healthc Qual.* 2003;25(2):4–10; quiz 10–11.
92. Kirkman MS, Williams SR, Caffrey HH, Marrero DG. Impact of a program to improve adherence to diabetes guidelines by primary care physicians. *Diabetes Care.* 2002;25(11):1946–1951.
93. Lenz ER, Mundinger MO, Hopkins SC, Lin SX, Smolowitz JL. Diabetes care processes and outcomes in patients treated by nurse practitioners or physicians. *Diabetes Educ.* 2002;28(4):590–598.
94. Mottur-Pilson C, Snow V, Bartlett K. Physician explanations for failing to comply with “best practices.” *Eff Clin Pract.* 2001;4(5):207–213.
95. Symons AB, Mahoney MC, Englert J, Mirand AL. Variations in approaches to breast cancer screening among primary care physicians. *J Cancer Educ.* 2002;17(4):205–210.
96. Representative articles by Weisman on the topic of benchmarking include:
 - Weisman NW, et al. Improving quality improvement using achievable benchmarks for physician feedback: A randomized controlled trial. *JAMA.* 2001;285:2871–2879.
 - Weisman NW, Allison JJ, et. al. Achievable benchmarks of care: The ABC’s of benchmarking. *J Eval Clin Pract.* 1999;5(3):2369–281.
 - Weisman NW, Allison JJ, Keife CI. Can data-driven benchmarks be used to set the goals of Healthy People 2010? *Am J Public Health.* 1999;89:1:61–65.
97. Baldrige National Quality Program. Malcolm Baldrige National Quality Award Recipients. Gaithersburg, MD: National Institute of Standards and Technology; Apr 15, 2003.
98. American Society for Quality. *Quality Management Systems—Guidelines for Process Improvements in Healthcare Organizations.* Issue 1: Nov 2000.
99. Shortell SM. High-performing healthcare organizations: Guidelines for the pursuit of excellence. *Hosp Health Serv Admin.* July/Aug 1985:7–35.
100. Shortell SM, O’Brien JL, et al. Assessing the impact of continuous quality improvement/total quality management: Concept versus implementation. *Health Serv Res.* 1995;30(2):377–401.
101. Shortell SM, Levin DZ, et al. Assessing the evidence on CQI: Is the glass half empty or half full? *Hosp Health Serv Admin Special CQI Issue.* 1995;40(1)4–24.
102. Shortell SM, Bennett CL, Byck GR. Assessing the impact of continuous quality improvement on clinical practice: What it will take to accelerate progress. *Milbank Q.* 1998;76(4):593–625.

103. Berwick DM, James B, Coyle MJ. Connections between quality measurement and improvement. *Med Care*. 2003;41:S130–S138.
104. Scott J, Schenkman M, et al. Exploring nursing home staff perceptions of communication and leadership to facilitate quality improvement. Unpublished. 2003.
105. Aiken LH, Sochalski J. Studying outcomes of organizational change in health services. *Med Care*. 1997;35(11):NF 6–18.
106. Levesque DA, Prochaska JM, et al. Organizational stages and processes for change for continuous quality improvement in health care. *Consult Psych J: Pract Res*. 2001;53(3):139–153.
107. Perspectives: shared visions-view pathways. *Joint Commission Newsletter*. Oct 2002.
108. Archer N. E-mail From Arkansas Foundation for Medical Care. Nov 17, 2003.
109. Burgess P. E-mail From Alabama Quality Improvement & Communications. Nov 17, 2003.
110. Jones J. E-mail from Delmarva. Nov 19, 2003.
111. Leyden T. E-mail from MPRO. Nov 17, 2003.
112. Lundblad JP. E-mail from Stratis Health. Nov 24, 2003.
113. Powell E. E-mail from Medical Review of North Carolina. Nov 19, 2003.
114. Royston M. E-mail from Qualis Health. Nov 17, 2003.
115. Stanley C. E-mail from Virginia Health Quality Center. Nov 18, 2003.
116. Weddle J. E-mail from Qsource. Nov 20, 2003.
117. Scott J, Vojir C, Jones K, Moore L. Critical Elements of Nursing Home’s Capacity to Create and Sustain Clinical Improvement. Unpublished research paper; 2003.
118. Berlowitz DR, Brandeis GH, Morris JN, et al. Deriving a risk-adjustment model for pressure ulcer development using the Minimum Data Set. *J Am Geriatr Soc*. 2001;49(7)866–71.

A. Study Protocol for Defining High Performing Health Care Providers

I. Introduction and Background

HSAG is conducting a CMS Special Study on *Identification and Synthesis of Components Essential to Achieving “High Performer” Status in Various Provider Types*. As part of this study, HSAG will conduct a literature review and gather information from other sources in order to provide CMS with a summary of what is known about high performance/high performers in selected health care settings (hospitals, nursing homes, home health agencies, and physician offices) and what has been published on measuring and comparing institutional performance. Reviews are to be restricted, as much as possible, to the organization and delivery of health care to Medicare beneficiaries, and to the clinical conditions in the 7th SoW and proposed 8th SoW. However, it is often the case that pertinent literature or other insight comes from sources not specifically focused on Medicare.

In completing the task of identifying dimensions of high performance, non-literature sources of information will also be reviewed. These include business-oriented standardization groups like the National Institute of Standards and Technology’s Malcolm Baldrige Award, accrediting bodies such as the Joint Commission on Accreditation of Healthcare Organizations (JCAHO), and programs such as CMS Bridges to Excellence. Finally, all QIOs that have a formal “recognition/awards program” for collaborators will be queried to ascertain their measurement strategies of performance.

II. Purpose of the Study

There are two study purposes. First, a literature review of statistical methods used by health services researchers to measure institutional high performance will be used to refine the working definition of high performance, develop measurement methods, and develop the high performance algorithm, Study Task 3. Second, an abstract review and non-literature source review will identify key quality characteristics of performance measurement in the four selected settings and their respective conditions from the 7th SoW and proposed 8th SoW.

III. Review Questions

Review questions are focused on the study purposes of identifying measurement methods for institutional performance and identifying the key quality characteristics of performance measurement in the selected health care settings.

A. Performance Measurement

What statistical methods, based on existing quality indicators, have been used for profiling and ranking institutions, and for classifying their performance? Do these methods discriminate among institutions? Are they meaningful to decision makers?

B. Key Clinical Indicators of Quality Performance

To what extent have clinical conditions and related performance measures from the current 7th Scope of Work and from the proposed 8th Scope of Work been the subjects of published research identifying high performance in hospitals, nursing homes, home care agencies, and physician offices?

IV. Review Protocol

The focus of the information search is on four health care settings and managed care organizations: hospitals, nursing homes, home health agencies, and physician offices. Separate protocols have been developed for reviewing performance measurement and for reviewing key quality characteristics.

A. Literature Review

1. Performance Measurement

a. Search strategy

The search strategy will focus on recent efforts to profile and compare health care institutional performance using statistical methods or models. We will use a combination of literature review, unpublished reports available to the investigators, and referenced citations found in review articles. The online literature search will be limited to articles published in the years 1997 to 2003.

For the online literature search, Endnote or Reference Manager software will be used to access the National Library of Medicine (NLM), PubMed, using combinations of MeSH major topics and keywords:

- ◆ hospital
- ◆ nursing homes
- ◆ home care
- ◆ physicians
- ◆ quality indicators, health care
- ◆ statistics & numerical data
- ◆ benchmarking
- ◆ health services research
- ◆ models, statistical

b. Article selection criteria

The aim of the study is to identify those articles related to measuring, profiling, and comparing organizational high performance in selected health care settings. Only articles written in English will be considered. A listing of other inclusion and exclusion criteria is found in Table A-1.

Table A-1. Selection Criteria and Inclusion and Exclusion Criteria for Literature Review		
Selection Criteria	Inclusion Criteria	Exclusion Criteria
Setting	Hospitals, nursing homes, home health agencies, physician offices	All other settings
Population	Adult Medicare and general	Special populations, such as pediatrics, medically underserved
Statistical methods and models	Institution as unit of analysis	Utilization, financial, manpower, clinical studies

c. Data extraction process

Article selection is a multi-stage process. In the first step, the article titles and abstracts are generated from the NLM database using the selected MeSH major topics. Using a data abstraction tool (Figure A-1), articles not meeting the inclusion criteria will be excluded by abstract review. The abstracts identified as being potentially relevant will be provisionally included for consideration. After retrieving the full texts of all potentially relevant citations, reviewers will assess the information to determine if criteria have been met or not. Citations initially included may be excluded at this stage. The process will be evaluated in the initial stages of the review, with more explicit criteria developed to improve the selection of articles for full review as needed.

A database will be developed for data entry and data aggregation. A summary of excluded articles will be appended to the final report.

To maintain consistency in the review process, we will use four HSAG reviewers including the Chief Quality Improvement Officer and the Vice President of Clinical Quality Assessment. As an initial evaluation, the four reviewers will abstract two to four of the same articles and compare results.

2. Key Clinical Indicators of Quality Performance*

a. Search strategy

The search strategy will initially focus on identifying published literature reviews/meta analyses related to quality improvement and high performance. For the online literature search, Endnote or Reference Manager will be used to access the NLM PubMed. Searches, limited to articles published in the years 2000 to 2003, will be conducted using combinations of MeSH major topics and keywords:

- hospitals
- nursing homes
- home care
- physicians
- Medicare
- heart failure, congestive
- surgical wound infection
- diabetes mellitus
- mammography
- pain
- urinary tract infections
- dyspnea
- confusion
- patient readmission
- rehabilitation of speech and

* This section was titled “Key Quality Characteristics” in the study protocol previously approved by CMS.

- Professional Review Organization
- literature review
- meta analysis
- quality assurance, health care
- quality indicators, health care
- pneumonia
- myocardial infarction
- immunization
- confusion
- walking
- decubitus ulcer
- restraint, physical
- infection control
- activities of daily living
- language
- anxiety
- behavioral symptoms
- diabetes screening
- hemoglobin A, glycosylated
- lipid profile
- breast cancer screening
- pneumococcal immunization

b. Abstract selection criteria

In identifying the key quality characteristics, the intent is not to conduct an academic and exhaustive review. The intent is to identify the recent work of health care leaders in measuring and defining performance that helps to characterize organizational high performer/high performance in selected health care settings for the conditions from the 7th SoW and proposed 8th SoW. Abstracts based on organizations outside of the United States, Canada, or the United Kingdom will be excluded from review. Only articles written in English will be considered. Additional inclusion and exclusion criteria are found in Table A-2.

c. Data extraction process

Abstracts generated from the NLM database will be reviewed for exclusionary criteria. Using the data abstraction tool (Figure A-1), abstracts found to be relevant will be reviewed for: study purpose, condition studied, performance measure studied, and principal findings. The information will be entered into a data base. A summary of excluded abstracts will be included in the final report as an appendix.

To maintain consistency in the review process, we will use four HSAG reviewers including the Chief Quality Improvement Officer and the Vice President of Clinical Quality Assessment.

Table A-2. Selection Criteria and Inclusion and Exclusion Criteria for Abstract Review

Selection Criteria	Inclusion Criteria	Exclusion Criteria
Setting	Hospitals (acute care, ambulatory care, intensive care unit, emergency department), nursing homes, home health agencies, physician offices	All other settings
Condition	CMS clinical topic areas specific to health care setting: hospital (pneumonia, myocardial infarction, heart failure, surgical wound infection); ambulatory care (diabetes eye exam, screening mammography, adult influenza and pneumococcal immunization); nursing home (pain, confusion, walking improvement, pressure sore, physical restraint, infection control); home	All other conditions

Table A-2. Selection Criteria and Inclusion and Exclusion Criteria for Abstract Review		
Selection Criteria	Inclusion Criteria	Exclusion Criteria
	health (activities of daily living, urinary tract infections, dyspnea, confusion, patient readmission, rehabilitation of speech and language, anxiety, behavioral symptoms)	
Population	Adult Medicare and general	Special populations such as pediatrics, medically underserved

B. Non-Literature Sources to Identify Key Quality Characteristics of Performance Measures

Non-literature sources may yield information on how others have defined or operationalized high performance. These include accrediting organizations such as NCQA, JACHO, and Malcom Baldrige. We will focus on describing the emphasis each organization places on administrative policies and procedures and performance indicators in its assessment of organizational performance.

CMS-contracted QIOs that have implemented recognition and awards programs will be surveyed to identify the common criteria used in selecting high performing organizations. Publications or reviews of QIOs on quality indicators or reports will also be reviewed.

C. Information Synthesis

We have proposed a definition of high performance using available provider indicators. Information on measurement methods of organizational performance will be used to refine this definition and to develop a measurement metric. A high performer algorithm will then be developed for identifying high performing institutions in four states: Arizona, New York, Connecticut, and Florida. Information on the institutional key quality characteristics will entail describing the reported measures of “high performer status” in the hospitals, nursing homes, home health agencies, and provider offices for the targeted conditions.

Figure A-1. Data Abstraction Tool



Reviewer Name: _____

Date of Review: _____

Exclusion criteria:

Applicable to both sets of articles:

- _____ abstract not available
- _____ article not available
- _____ study population
- _____ language

Dimensions of performance:

- _____ healthcare setting
- _____ condition
- _____ not US, UK, or CAN
- _____
- _____
- _____

Methods:

- _____ utilization study
- _____ manpower study
- _____ clinical study
- _____ not statistical methods
- _____
- _____

CITATION (skip if all info is attached):

Author(s): _____

Title: _____

Published In: _____

Editor (if any): _____

Pub. Date: _____ Volume: _____ Number: _____

Pages: _____ Edition: _____ Circle one: Journal

Publisher: _____ City: _____ URL: _____

Keywords:	1
	2
	3
	4
	5

Book Dissert. Unpub'd

PLEASE ATTACH CITATION AND ABSTRACT

APPENDIX A: STUDY PROTOCOL FOR DEFINING HIGH PERFORMING HEALTH CARE PROVIDERS

1. Study Design (Check one):

- RCT (Randomized Controlled Trial)
- CCT (Clinical Controlled Trial)
- CBA (Controlled Before & After Study)
- ITS (Intermittent Time Series)
- Simple Pre/Post Study
- Survey
- Qualitative (Case Study, Observational)
- Not a study (Book, Editorial, Opinion, Theory)
- Other _____

2. Study Setting (Check all that apply):

- | | | |
|---------------------------------------|----------------------------------|--------------------------------------|
| <i>Hospital</i> | <i>Physician</i> | <input type="checkbox"/> Local |
| <input type="checkbox"/> in-pt | <input type="checkbox"/> group | <input type="checkbox"/> State |
| <input type="checkbox"/> out-pt | <input type="checkbox"/> solo | <input type="checkbox"/> Multi-state |
| <input type="checkbox"/> ICU | <input type="checkbox"/> staff | <input type="checkbox"/> National |
| <input type="checkbox"/> ER | <input type="checkbox"/> primary | |
| <input type="checkbox"/> Home care | <input type="checkbox"/> spec | |
| <input type="checkbox"/> Nursing home | | |
| <input type="checkbox"/> Other: _____ | | |

3. Primary Purpose/Focus:

- Statistical methods (go to # 6 below)
- Benchmarking
- Determinants of quality/performance
- Governance / leadership
- Measurement (Indicator dev/testing)
- Physician performance
- QI Study
- Theoretical
- Use of technology/information systems
- Other: _____

4. Condition studied:

- | | | |
|--|--|--------------------------------------|
| <i>Hospital:</i> | <i>Home Care:</i> | <i>Physician:</i> |
| <input type="checkbox"/> pneumonia | <input type="checkbox"/> ADLs | <input type="checkbox"/> retinopathy |
| <input type="checkbox"/> heart failure | <input type="checkbox"/> U.T.I | <input type="checkbox"/> HgbA1c |
| <input type="checkbox"/> myocardial infarction | <input type="checkbox"/> dyspnea | <input type="checkbox"/> lipids |
| <input type="checkbox"/> surg wound inf | <input type="checkbox"/> confusion | <input type="checkbox"/> mammo |
| <i>Nursing Home:</i> | <input type="checkbox"/> re-hosp | <input type="checkbox"/> PPV |
| <input type="checkbox"/> decubitus ulcer | <input type="checkbox"/> speech/language | |
| <input type="checkbox"/> ADLs | <input type="checkbox"/> anxiety | |
| <input type="checkbox"/> confusion | <input type="checkbox"/> behavior sx. | |
| <input type="checkbox"/> infection control | | |
| <input type="checkbox"/> pain | | |
| <input type="checkbox"/> restraints, physical | | |
| <input type="checkbox"/> walking | | |

5. Dimension of Performance studied:

- | | |
|--|--|
| <input type="checkbox"/> access | <input type="checkbox"/> hosp pt flow |
| <input type="checkbox"/> availability | <input type="checkbox"/> productivity/efficiency |
| <input type="checkbox"/> pt sat/complaints | <input type="checkbox"/> equality of care |
| <input type="checkbox"/> structural measures | <input type="checkbox"/> medical error/pt safety |
| <input type="checkbox"/> process measures | <input type="checkbox"/> Other: _____ |
| <input type="checkbox"/> outcome measures | |

(For abstract review, go to #7.)

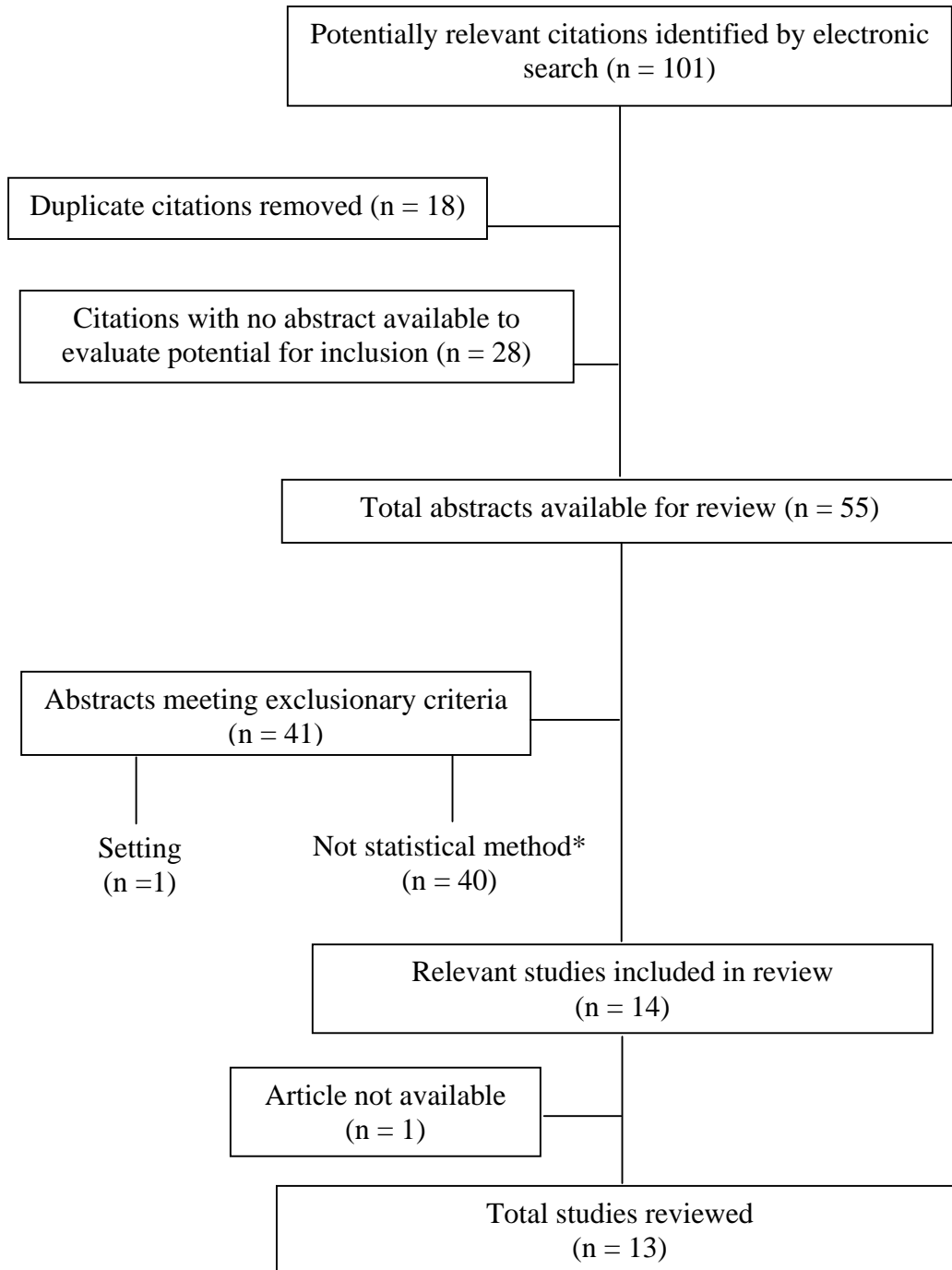
6. Statistical method / model used to identify institutional high performance

7. Principal Findings:

8. Study Limitations:

B. Results of MeSH Search on Performance Methods

Figure B-1. MeSH Search Performance Methods



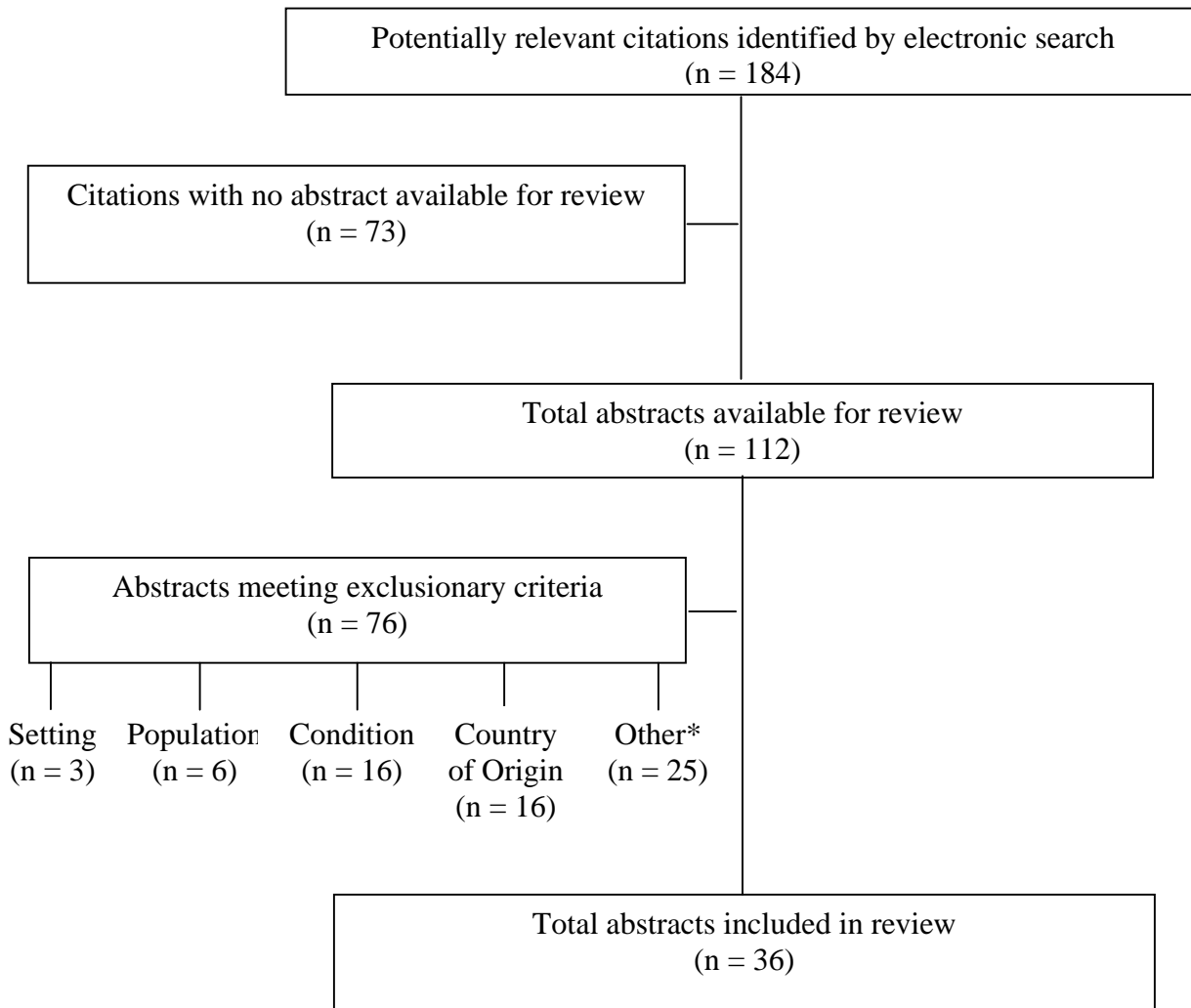
* Includes utilization, manpower, economic, clinical, and administrative data collection studies.

Table B-1. Results of MeSH Search on Performance Methods								
Performance Methods: Total Citations Found, Excluded, and Articles Reviewed								
MeSH Major Topic Search Terms	Total Citations	Criteria for Exclusion						Articles Reviewed
		No Abstract Available	Duplicates	Population or Language	Setting	Not Statistical Methods ¹	Article not Available	
Setting + benchmarking, quality indicators, statistics, and numerical data	101	28	18	0	1	40	1	13

¹ Includes utilization, manpower, economic, clinical, and administrative data collection studies.

C. Results of MeSH Search on Key Clinical Indicators of Quality Performance

Figure C-1. Key Clinical Indicators of Quality Performance



* Studies were excluded if they focused on utilization or cost studies, clinical guideline development, and data collection methods.

Table C-1. Results of MeSH Search on Key Quality Characteristics of Performance

Key Quality Characteristics of Performance: Total Citations Found, Excluded, and Abstracts Reviewed by MeSH Major Topic Search Terms								
MeSH Major Topic Search Terms	Total Citations	Criteria for Exclusion						Articles Reviewed
		No Abstract Available	Setting	Population	Condition	Country of Origin	Other ¹	
Hospital + conditions	43	12	0	1	0	4	9	18
Nursing home + conditions	34	10	0	0	10	3	0	10
Homecare + conditions	6	1	0	0	3	1	0	1
Physician + conditions	33	8	0	4	9	8	0	4
Medicare/PRO + conditions	68	42	3	1	4	0	15	3
Total	184	73	3	6	26	16	24	36

¹ Other studies were excluded because they focused on utilization or cost studies, clinical guideline development, and data collection methods.