Environmental Scan of Public Reporting Programs and Analysis

Final Report: Executive Summary

September 23, 2010

So O’Neil
John Schurrer
Sam Simon

MATHEMATICA
Policy Research, Inc.

Contract Number:
NQF2009-07

Mathematica Reference Number:
06738.500

Submitted to:
National Priorities Partnership
National Quality Forum
601 Thirteenth Street, NW
Suite 500N
Washington, DC 20005
Project Officer: Christy Olenik

Submitted by:
Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: So O’Neil

Environmental Scan of Public Reporting Programs and Analysis

Final Report: Executive Summary

September 23, 2010

So O’Neil
John Schurrer
Sam Simon

MATHEMATICA
Policy Research, Inc.

EXECUTIVE SUMMARY

The many different public reporting initiatives that aim to promote quality and efficiency in the health care system speak to a growing momentum in support of transparency in health care and evidence-based improvement. At the same time, the variety of purposes, audiences, and data sources associated with such initiatives can make it difficult to identify opportunities for coordination in pursuit of a national agenda for quality improvement.

To help identify potential areas for coordination in public reporting, the National Quality Forum (NQF) contracted with Mathematica Policy Research to assess the current landscape of public reporting in health care. The assessment also included the examination of public reporting in the context of a conceptual framework for understanding quality, as public reporting within such a framework can help develop a common understanding of quality in health care among stakeholders (Institute of Medicine 2001, 2006). To enhance and broaden understanding of quality in health care, Hibbard and Pawlson (2004) have promoted the development and use of a framework that is consistent with the Institute of Medicine's six aims for the health care system: ensuring that care is safe, timely, effective, efficient, equitable, and patient centered. The NQF has subsequently developed an integrated framework incorporating these six aims and reflecting national priorities; this integrated framework was reviewed and endorsed through its multistakeholder consensus process. The key elements of the framework, adapted to include the measurement domains used in Mathematica's analysis, are depicted in Figure I.

Figure I. Domains of the NQF Integrated Framework

Phases of Care

- Population at Risk
- Initial Evaluation and Management
- Follow-Up Care

National Priority Areas

- Patient and Family Engagement
- Population Health
- Safety
- Care Coordination
- Palliative and End-of-Life Care
- Overuse

Measurement Domains

- Access
- Process
- Outcome
- Cost
- Structure

Note: Phases of care are from the NQF patient-focused episode-of-care measurement framework, which includes the following components: population at risk (health promotion, primary and secondary prevention); initial evaluation and management (onset of clinical illnesses and initial assessment); and follow-up care (coordination and transitional phase) (NQF 2009).

National priority areas were put forward by the National Priorities Partnership, which is convened by NQF and represents a diverse range of high-impact stakeholder organizations focused on health care improvement. The priority areas in the figure described in further detail are (1) engage patients and families in managing their health and making decisions about their care; (2) improve the health of the population; (3) improve the safety and reliability of America's health care system; (4) ensure that patients receive well-coordinated care within and across all health care organizations, settings, and levels of care; (5) guarantee appropriate and compassionate care for patients with life-limiting illnesses; and (6) eliminate overuse while ensuring the delivery of appropriate care.

The objectives of the environmental scan and analysis were to provide an overview of current public reporting efforts, to identify opportunities for harmonization among programs and publicly reported measures, and to identify gaps in measurement in the context of the NQF's integrated framework. Such information can contribute to current knowledge about public reporting in health care and can inform decision making related to public reporting, including the development of a standardized community dashboard of core quality measures. The objectives of the project centered on four research questions:

1. What domains and measures are captured in public reporting activities?
2. How do publicly reported measures map to, and converge with, the NQF integrated framework?
3. How do publicly reported measures diverge from the integrated framework? What are the gaps in reporting?
4. Among the NQF integrated framework areas addressed, what is the convergence (or congruity) between publicly reported measures within specific domains of the integrated framework? What is the divergence between measures within specific domains of the integrated framework?

Methods for Conducting Environmental Scan and Analysis of Public Reporting Programs

From February to September 2010, Mathematica conducted the environmental scan and analysis of public reporting programs, based on a sample of programs identified in collaboration with NQF. Programs in the initial sample included those identified in previous research by Roski and Kim (2009), Cronin and Shearer (2005), and the Ambulatory Care Quality Alliance (2010); and in the Agency for Healthcare Research and Quality's Report Card Compendium. Members of the following initiatives were also included: Aligning Forces for Quality, the Better Quality Information for Medicare Beneficiaries Pilot Project, Charter Value Exchanges, Network for Regional Healthcare Improvement, National Academy for State Health Policy, and the National Association of Health Data Organizations initiatives. Additional programs were identified by six key informants consulted for the study and through a general literature search.

Through this process, we identified 332 programs, of which 162 met project criteria for public reporting. We defined a public reporting program as one that has information publicly available through either web-based or paper documents without any associated fees. However, for this project, we included only programs with information available through the Internet. Proprietary programs were included, provided they made at least some information available to the general public without subscription. In consultation with our six key informants, we narrowed the sample to 72 programs, stratified by geographic area of focus and date of public reporting initiation, to ensure that it was representative of the larger universe of programs identified.

Information on the 72 identified public reporting programs and their quality measures were collected and entered into a Structured Query Language (SQL) database; SQL is a computer language designed for relational database management systems. The database included fields for data entry related to program and measure domains.

- ***Program domains*** included audience; availability of information; contact information; first published report date; frequency of reports; geographic level—national, state, county, community, and other (metropolitan statistical area, health service area, hospital referral region); most recent report date; payer type; program description; program name; program website; publication mode; report link; report name; sponsor; type of organization; and time of public reporting adoption.

- ***Measure domains*** included condition; data source; measure title, measurement domain; national priority area; NQF endorsement; phase of care; denominator; numerator; target population; and unit of analysis.

After cataloging the sampled program and measure information, we conducted four levels of analyses: (1) descriptive analysis of publicly reported programs and measures to provide an overview of public reporting, (2) mapping of unique measures to the NQF integrated framework to assess the types of measures reported in each domain of a framework for understanding quality, (3) analysis of convergence and divergence in public reporting to the integrated framework to provide greater depth of information about reporting within domains of a framework for understanding quality, and (4) analysis of congruity (convergence) among measures within specific domains of the framework to examine key factors affecting the potential for measure harmonization.

Descriptive Analysis of Programs and Measures

The 72 sampled programs reported a total of 4,254 measures, of which 1,685 represented unique (or unduplicated) measures. The total number of measures reported per program ranged from 1 to 261, with a mean of 59.

Characteristics of Sampled Programs

Although the public reporting programs reviewed exhibited a variety of characteristics, the "typical" program was a state-level initiative begun in the past five years, sponsored by a state agency, and directed at the general public for purposes of accountability or informed consumer choice. Key characteristics of all programs reviewed are summarized next.

- ***Date of initiation.*** More than 70 percent of programs selected for review began public reporting in 2005 or afterward.

- ***Geographic scope.*** Most programs reviewed were state-level reporting programs (64 percent). Programs national in scope were the next most common (15 percent), followed by regional programs (10 percent), county-level reporting programs (7 percent), and community-level programs (4 percent).

- ***Organizational sponsorship.*** State agencies sponsored the largest proportion of programs (33 percent), followed by multistakeholder organizations (24 percent), consumer/advocacy groups (13 percent), employer business groups (10 percent), and the federal government (1 percent). A mix of other organizations (for example, academic institutions, commercial health plans, hospital associations, and provider groups) sponsored the remaining 19 percent of programs.

- ***Target audience.*** Most public reporting programs targeted consumers or the general public (88 percent). Other identified audiences included health care providers/managers (11 percent), purchasers/benefits designers (11 percent), payers (4 percent), and policymakers/regulators (4 percent).

- ***Purpose.*** Most of the selected programs reported quality measures for the purpose of informing consumer choice (92 percent) and public accountability (90 percent). The selected programs less often reported measures for purposes of quality improvement (33 percent), accreditation/certification (4 percent), or for payment incentive (3 percent).

Characteristics of Quality Measures Used in Public Reporting Programs

The measures used by the sampled public reporting programs reflected several different measurement domains. However, the typical program reported facility-specific, NQF-endorsed, hospital quality measures based on administrative claims data from all payers, for chronic cardiovascular and pulmonary conditions affecting people older than 65. Characteristics of the quality measures used across all reporting programs reviewed are summarized next.

- ***Sources of data.*** Most programs (85 percent) relied on administrative claims as a key data source. Patient surveys (65 percent) were the next most common source, followed by facility surveys (28 percent).
- ***Payer type.*** Seventy-two percent of programs reported measures relevant to populations covered by all payers. Other programs included data from one or more of the following payer types: commercial (19 percent); Medicare (10 percent); and Medicaid (1 percent). Three percent of programs did not specify a payer type.
- ***Unit of analysis.*** Eighty-one percent of programs used facility, such as hospital or nursing home, as the unit of analysis. Group practices were the next most common unit of analysis (38 percent), followed by health plan (28 percent), and individual practice (13 percent).
- ***Setting of care.*** Inpatient hospitals were the most common setting of care for measure reporting (74 percent), followed by clinicians' office (58 percent) and nursing care facility (26 percent). Other settings of care were reported by fewer than 15 percent of programs.
- ***Age groups.*** The largest number of programs (more than 75 percent) reported measures specific to people older than 65. Measures specific to the general adult population were the next most common (69 percent of programs), but slightly more than half of the programs (51 percent) also reported measures specific to children.
- ***Conditions.*** Seventy-eight percent of programs reported measures related to chronic cardiovascular conditions, and 71 percent of programs reported measures related to chronic pulmonary conditions.
- ***Disparities.*** Relatively few of the sampled programs reported measures related to disparities in care, with 6 percent addressing racial/ethnic disparities, 4 percent addressing socioeconomic disparities, and 6 percent addressing some other type of disparity.
- ***NQF endorsement.*** Seventy-eight percent of programs used NQF-endorsed measures.

Mapping and Analysis to Assess Convergence and Divergence with a Framework for Understanding Quality

To assess the extent to which public reporting programs align with a quality framework reflective of national priorities, we mapped measures to the NQF integrated framework and examined convergence with or divergence from it. We used both total (including duplicated) measures and unique (unduplicated) measures in the analysis. We defined convergence in terms of the number of programs reporting and the number of measures being reported that corresponded to a specific domain. Divergence from the framework or gaps in public reporting are defined as domains in the integrated framework in which few programs are reporting and few measures are being reported. The key domains of the NQF integrated framework assessed for convergence and divergence included the three phases of care, the six national priority areas, and the five measurement domains shown in Figure I.

Our analysis of programs and measures indicates that the degree of convergence varies considerably across domains. Table I provides an overview of the percentage of programs, percentage of duplicated measures, percentage of unique measures, and mean number of measures reported within each domain of the integrated framework. We highlight key findings in the bullets below.

Table I. Domains Captured by Programs and Measures

	Programs	Duplicated Measures	Unique Measures	Mean
Total	72	4,254	1,685	59
	% of Total			# of Measures
Phase of Care				
Population at risk	71	12	8	10
Initial evaluation and management	92	38	40	24
Follow-up care	96	47	46	29
Not classified	78	27	35	20
National Priority Area				
Patient and family engagement	72	11	12	9
Population health	68	13	8	12
Safety	71	16	16	13
Care coordination	83	32	22	23
Palliative and end-of-life care	15	4	7	17
Overuse	60	10	13	10
Measure not classified	74	16	24	13
Measurement Domain				
Access	54	2	3	3
Cost and utilization	82	26	32	19
Structure	47	6	34	7
Process	85	39	26	27
Outcome	90	33	11	22

Notes: Not all categories are mutually exclusive, and programs may have measures in more than one category. Therefore, column percentages may add up to more than 100 percent.

For duplicated measures, several programs may report the same measure, in which case the measure is counted once for each time it is reported. Thus, if two programs report the same measure, the measure is counted twice.

A unique measure is defined as having the same measure description, measurement domain, national priority, and phase of care. Multiple programs may report the same measure, but in our analysis of unique measures, this measure is only counted once.

Mean is calculated among programs that reported measures in the specific category. Programs that did not report any measures in the category were not included in the calculations.

Phases of Care

Initial evaluation and management and follow-up care were the phases of care with the highest level of convergence. More than 90 percent of programs reported measures associated with these two phases, with 38 percent of measures mapping to initial evaluation and management and 47 percent mapping to follow-up care. Among programs that report in these two phases, the average number of measures reported per program was also high; programs reported an average of 24 initial evaluation and management measures and 29 follow-up care measures. Examination of unique measures also reflects this pattern, as 40 and 46 percent of unique measures are initial evaluation and management, and follow-up care, respectively.

The phase of care with the lowest percentage of programs reporting was the population at risk phase. Seventy-one percent of programs reported one or more measures that could be mapped to this area, and 27 percent of duplicated measures and 8 percent of unique measures mapped to this area. The average number of measures among programs reporting in the population at risk phase was also relatively lower compared with that for programs reporting measures in the initial management and evaluation (24 measures) and follow-up care (29 measures) phases; programs reported an average of 10 population at risk measures.

National Priorities

Care coordination was the priority area that had the highest level of convergence. Compared with other priority areas, most programs reported measures in the care coordination area (83 percent); the highest percentage of measures could be mapped to this priority area (32 percent), and the highest average of number measures per program was associated with this area (23 measures per program). Reflecting this pattern, the largest proportion of unique measures mapped to the care coordination priority area (22 percent).

Overuse measures were reported by more than half of reporting programs. Sixty percent of programs reported overuse measures, with a mean of 10 measures per program (out of 59) mapping to this domain. Overuse measures also accounted for 13 percent of all unique measures reported.

A majority of programs reported population health measures, but the number of measures they reported was low. Though 68 percent of programs reported at least one population health measure, only 13 percent of duplicate measures and 8 percent of unique measures mapped to the population health priority area.

Few programs reported palliative and end-of-life care measures. Fifteen percent of programs reported palliative and end-of-life care measures, and 4 percent of unique measures could be mapped to this priority area. Among unique measures, only 7 percent mapped to palliative and end-of-life care measures. However, among programs that reported within this area, the average number of measures reported per program was relatively high (17 measures).

Measurement Domain

Public reporting among sampled programs was highly convergent with measurement domains of outcome, process of care, and cost and utilization. Eighty-five percent of programs reported process of care measures, 90 percent reported outcome measures, and 82 percent reported cost and utilization measures. The percentage of measures associated with each area was 39 percent for process of care, 33 percent for outcome, and 26 percent for cost and utilization. The average number of measures reported per program in each of these measurement domains was also relatively high, with 27 measures (process), 22 measures (outcome), and 19 measures (cost and utilization) reported. Examination of unique measure also reflects this pattern, where 26 percent are process measures, 34 percent are outcome measures, and 32 percent are cost and utilization measures.[1]

Access and structure measures were the least reported measures. Only 54 and 47 percent of programs reported access and structure measures, respectively. Access measures included those related to timely access to care and services; structure measures included those related to supports for the provision of health care (for example, facility amenities, workforce hours, and availability of health information technology). In addition, only 2 percent of measures could be mapped to the access domain and 6 percent mapped to the structure domain. Among programs reporting access and structure measures, the average number of measures reported per program was three and seven measures, respectively. Although access and structure measures made up a larger percentage of unique measures (3 and 11 percent, respectively), there was still considerably lower reporting within these domains in comparison with other measurement domains.

Convergence and Divergence Within a Specific Domain of a Framework for Understanding Quality

After assessing the extent to which public reporting maps to a quality framework, our next step was to understand the degree to which it is possible to compare quality across reporting programs. To assess comparability, we conducted an analysis of convergence and divergence of measures within a specific domain of the NQF integrated framework. For the study, we selected two areas of measurement—cholesterol management and heart failure quality of care—to illustrate the process for conducting such an analysis. These two areas were selected because they had several measures that fell into specific domains across the integrated framework: follow-up care (phase of care), care coordination (national priority), and clinical processes of care (measurement domain). For these analyses, we used *convergence* to indicate similarities along key characteristics of measures and *divergence* to indicate differences along these characteristics. The measure characteristics along which we compared measures included measure description, numerator, and denominator; purpose of measurement; data source; target population; geographic level of reporting; use of NQF-endorsed measures; unit of analysis; and setting of care.

Eight cholesterol management and six heart failure quality-of-care measures were identified and analyzed. Overall, the eight cholesterol management measures showed a fair amount of alignment in purpose of measurement, data source, geographic level of program, use of NQF-endorsed measures, and unit of analysis. Similarly, the purposes, data sources, NQF endorsement, unit of analysis (facility), and setting of care (inpatient) were the same for the six heart failure quality-of-care measures. Within both areas of measurement, however, the numerator and denominator differed in instrumental ways, which would make valid comparisons between the measures within an area difficult.

---

[1] Outcome measures were mainly patient safety and outcome measures (55 percent of unique measures), but also included readmission, morbidity, mortality, health-related quality of life, intermediate outcome, functional status measures, and other outcomes. Cost measures included those related to procedure utilization, an episode of care, length of stay, hospital readmission, imaging, per capita costs, emergency department visits, medication prescribing, other service costs, and other cost and resource use. Process measures include those related to clinical care processes, healthy lifestyle behaviors, care coordination, patient and family engagement, prevention services, safety practices, and other processes.

Implications for Public Reporting in Health Care

Findings from the mapping and analysis of public reporting programs and their measures have several implications for a national quality agenda.

Our analysis suggests that although there is a high degree of convergence between reporting initiatives and several domains of a quality framework reflective of national priorities, relatively less attention has been paid to public reporting of population-based measures (including measures of disparity), public reporting within the overuse and population health national priority areas, and public reporting within the measurement domain of access to care. Given the importance of these issues to the current policy agenda, these might be areas to promote in public reporting. They might also provide opportunities around which to engage consumers and other stakeholders in public dialogue—especially among the state and federal government-sponsored programs that appear to dominate the reporting landscape.

Moreover, as stakeholders have an interest in better coordinating public reporting efforts, our analysis suggests that further effort might be needed to harmonize reporting, through the development of standardized measurement specifications. The sample of programs reviewed for this study showed that the number and types of measures reported vary considerably across programs and among measures within the same topic. Although such variation is expected given differences in resources, purpose, and audience, the ability to make valid comparisons is especially critical to the development of measures that can inform public policy at a national level.

Limitations of the Study

Based on our methods, several caveats should be considered in interpreting findings. First, the sample was selected to represent the diversity of public reporting programs according to key informant input and other study criteria (for example, definition of public reporting program, geographic representation, and period of public reporting initiation). Therefore, the sample might not represent the universe of programs. Second, decisions regarding the categorization of programs and measures were subject to team interpretation of definitions and guidance. However, several procedures were undertaken to ensure internal consistency of the data, including routine and comprehensive quality checks and standard training and procedures for data entry. Finally, information cataloged was limited to information available through the public reporting program's website, and how measures were cataloged was subject to how they were presented on the website.

Conclusion

Our analysis suggests that if current patterns persist there will continue to be considerable variation in measurement and reporting. This variation in practice may well contribute to innovation in this evolving field. However, it also creates challenges to efforts to develop a coordinated national approach to quality and efficiency in health care. Continued assessment of potential areas for development and coordination of efforts will enhance the quality and usefulness of public reporting initiatives.

References

Ambulatory Care Quality Alliance."Environmental Scan: Physician and Other Clinician and Consumer Reports."Washington, DC: AQA, March 24, 2010.

Cronin, Carol, and Anna Shearer. "The State-of-the-Art of Online Hospital Public Reporting: A Review of Fifty-One Websites." Report prepared for the Centers for Medicare & Medicaid Services Three State Pilot Project. 2nd ed. Baltimore, MD: Delmarva Foundation, July 2005.

Hibbard, Judith, and L. Gregory Pawlson. "Why Not Give Consumers a Framework for Understanding Quality?" Joint Commission Journal on Quality and Safety, vol. 30, no. 6, June 2004, pp. 347–351.

Institute of Medicine. Crossing the Quality Chasm; A New Health System for the 21st Century. Washington, DC: National Academy Press, 2001.

Institute of Medicine. Performance Measurement; Accelerating Improvement. Washington, DC: National Academy Press, 2006.

National Quality Forum."Measurement Framework: Evaluating Efficiency Across Patient-Focused Episodes of Care." Washington, DC: NQF, 2009

Roski, Joachim, and Min Gayles Kim. "Current Efforts of Regional and National Performance Measurement Initiatives Around the United States." American Journal of Medical Quality (published electronically ahead of print), September 3, 2009.

MATHEMATICA
Policy Research, Inc.

www.mathematica-mpr.com

Improving public well-being by conducting high-quality, objective research and surveys

Princeton, NJ ■ Ann Arbor, MI ■ Cambridge, MA ■ Chicago, IL ■ Oakland, CA ■ Washington, DC