RAND HEALTH

The RAND Corporation is a nonprofit institution that helps improve policy and decisionmaking through research and analysis.

This electronic document was made available from www.rand.org as a public service of the RAND Corporation.

Skip all front matter: Jump to Page 1 ▼

Support RAND

Browse Reports & Bookstore

Make a charitable contribution

For More Information

Visit RAND at www.rand.org

Explore RAND Health

View document details

Limited Electronic Distribution Rights

An Evaluation of the Use of Performance Measures in Health Care

Cheryl L. Damberg, Melony E. Sorbero,

Susan L. Lovejoy, Katharine Lauderdale,

Samuel Wertheimer, Amber Smith,

Daniel Waxman, Christopher Schnyer

RAND HEALTH

The RAND Corporation is a nonprofit institution that helps improve policy and decisionmaking through research and analysis. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

Preface

The National Quality Forum (NQF), a private, nonprofit membership organization committed to improving health-care quality performance measurement and reporting, was awarded a contract with the U.S. Department of Health and Human Services (HHS) to establish a portfolio of quality and efficiency measures. The portfolio of measures would allow the federal government to examine how and whether health-care spending is achieving the best results for patients and taxpayers. As part of the scope of work under the HHS contract, NQF was required to conduct an independent evaluation of the uses of NQF-endorsed measures for the purposes of accountability (e.g., public reporting, payment, accreditation, certification) and quality improvement (QI). In September 2010, NQF entered into a contract with the RAND Corporation for RAND to serve as the independent evaluator.

This report presents the results of the evaluation study. It describes how performance measures are being used by a wide array of organizations and the types of measures being used for different purposes, summarizes key barriers and facilitators to the use of measures, and identifies opportunities for easing the use of performance measures moving forward.

This report will be of interest to private and public organizations that are engaged in performance measurement, including measure developers, organizations that work to promote the alignment and use of measures, NQF as the national measure-endorsement entity, and organizations that use performance measures to support improvements in care delivery.

This work was sponsored by HHS. The research was conducted in RAND Health, a division of the RAND Corporation. A profile of RAND Health, abstracts of its publications, and ordering information can be found at http://www.rand.org/health.

Contents

Figures

Tables

Summary

Over the past two decades, a wide array of standardized health-care performance measures have been developed, and a large number of these have been endorsed by the National Quality Forum (NQF). During this period, there was also substantial growth in the number of entities using health-care performance measures for a host of purposes. Quality and efficiency measures are now embedded throughout the U.S. health-care system, as demonstrated by their widespread use for accountability; accreditation, certification, credentialing, and licensure; public reporting; pay for reporting (P4R); pay for performance (P4P) and performance-based contracting; tiering and construction of narrow provider networks; quality improvement (QI); and public recognition. Although there has been significant growth in the use of performance measures (Madison, 2009; National Committee for Quality Assurance, 2010), there has been no systematic attempt to catalog the various ways in which measures are being used and what opportunities exist for enhancing the end users' ability to use performance measures to achieve their desired objectives.

Under requirements in the Medicare Improvements for Patients and Providers Act of 2008 (Pub. L. 110-275, 2008), the Secretary of the U.S. Department of Health and Human Services (HHS) contracted with NQF to carry out a set of activities related to the establishment of a portfolio of quality and efficiency measures that would enable the federal government to examine how and whether health-care spending is achieving the best results for patients and taxpayers. One component of this work was to engage an independent third party to evaluate the uses of performance measures for purposes of accountability and QI. A recently completed NQF study highlighted the need for understanding how measures are being used (Booz Allen Hamilton, 2010).

NQF engaged the RAND Corporation to conduct an independent examination of the use of performance measures, with particular interest in the use of NQF-endorsed measures. The goal was to better understand

- the current state of performance measure use across the broad spectrum of end-user types
- areas in which gaps in measures exist that hinder the end users' ability to apply measures to support the achievement of their desired goals
- how the larger measurement enterprise (i.e., measure developers, measure endorsers, foundations and government agencies that support measure development and implementation) might better support the use of performance measures.

The RAND team reviewed the peer-reviewed published literature and documents drawn from the non–peer-reviewed (i.e., gray) literature of the past five years as part of this project

and found no existing study that cataloged all uses or types of measures within each use. The absence of a comprehensive assessment underscores the importance of this project.

Study Questions and Approach

The overarching questions addressed in this study were as follows:

- How are performance measures being used in practice?
- What factors are influencing measure use, particularly endorsed measures?
- What types of system changes have occurred as a result of measure use?
- What improvement areas could facilitate the use of standardized endorsed performance measures by various end users?

Within the six-month time frame and scope of the project, the study team could not conduct an exhaustive review of the use of performance measures for various purposes. Instead, the project focused on conducting a scan of the current measure-use landscape to address the study questions.

The study design consisted of two data-collection methods: (1) interviews with end-user key informants (n = 30 end-user organizations) and (2) review of publicly available documents and materials from websites (n = 70 end-user organizations, of which 30 participated in the key-informant interviews). For both data-collection methods, we selected a purposive sample of organizations from 11 different categories of end-user organizations or entities (e.g., consumers, purchasers, community collaboratives) and end uses. The missions of the end users varied and included organizations focused on improving quality, creating information to help consumers select providers, working to assess the competencies of providers and health plans to deliver high-quality care, structuring incentives to drive improvements in care, and promoting competition on price and quality. We interviewed the executive director, president, or chief executive officer or the lead person responsible for quality-measurement activities within the organization.

We classified the use of performance measures into four broad categories of measure end use (i.e., the purpose for which measurement is used):

- QI
- public reporting
- accreditation, certification, credentialing, and licensure
- payment applications (e.g., financial incentives, tiering, narrow networks).

Organizations could be recorded as using measures for up to a maximum of four of these different end uses. We also recorded end uses that fell outside these four categories.

We focused our review on the primary end user of performance measures. The primary end user is defined as any organization that is directly engaged in the implementation of health-care quality or efficiency measures to assess the performance of providers of health care (i.e., hospitals, physicians, nursing homes, health plans). The primary end user is the organization that gathers the data; gains access to data for use in constructing measures or, in some

cases, uses preexisting measures; uses data to construct measures; computes the performance score of providers; and makes use of the resulting performance information.

Primary end users make performance information available with the expectation that others (i.e., providers, consumers) will use the performance measures to drive system changes, make behavior changes, or inform decisions. The others we define as secondary end users of the information. An example of the secondary use of performance measures would be a medical group changing its behavior in response to receiving performance scores from a community collaborative or consumers using comparative performance information on health plans to select a plan. The scope of this contract did not permit the exploration of secondary uses, although this would be an important area to address in a future impact study.

Key Findings

Use of Performance Measures

- According to our review of documents from end users, many organizations are using measures for multiple purposes, such as public reporting and P4P. There were a total of 127 uses documented for the 70 organizations included in the sample. Forty of the 70 organizations reported more than one use of performance measures.
  - During the key-informant interviews, several end users reported other uses of performance measures beyond the four broad categories of focus for this study, including evaluation (e.g., evaluating the impact of an intervention) and compliance monitoring of health plan contracts.
- Public reporting (64 percent [45 of 70]) and QI (69 percent [48 of 70]) were the most commonly reported uses among the four end uses (see Figure S.1). In contrast, uses for payment (33 percent [23 of 70]) and accreditation, certification, credentialing, or licensure (16 percent [11 of 70]) were reported less often.
- End users reported using measures drawn from seven core measurement domains: structure, process, outcome, access, safety, costs, and patient experience.[1] At this juncture, process measures are the most common type of measure being used (83 percent of measure uses report using process-of-care measures). The use of different types of measures varies by type of end use and the setting of care in which measures are being applied:
  - Structural measures are most likely to be used by those doing accreditation, certification, credentialing, or licensure (46 percent [five of 11] of these uses); cost measures are most likely to be used by those doing public reporting and P4P (49 percent [22 of 45] and 52 percent [12 of 23] of these uses, respectively); and patient-experience measures are most likely to be used by those doing public reporting (76 percent [34 of 45] of this use). (See Table D.1 in Appendix D.)
  - Access measures are the least commonly used type of measure overall, with 24 percent (31 of 127) of all uses reporting inclusion of this measurement domain. In our sample of organizations, access measures are reported being used in 18 percent (two of 11) of accreditation, certification, credentialing, or licensure uses; 13 percent (three of 23) of

---

[1] We assessed measure domains and endorsement status for each setting of care per measure use per organization.

Figure S.1
Frequency of Measure Uses (n = 70 organizations)
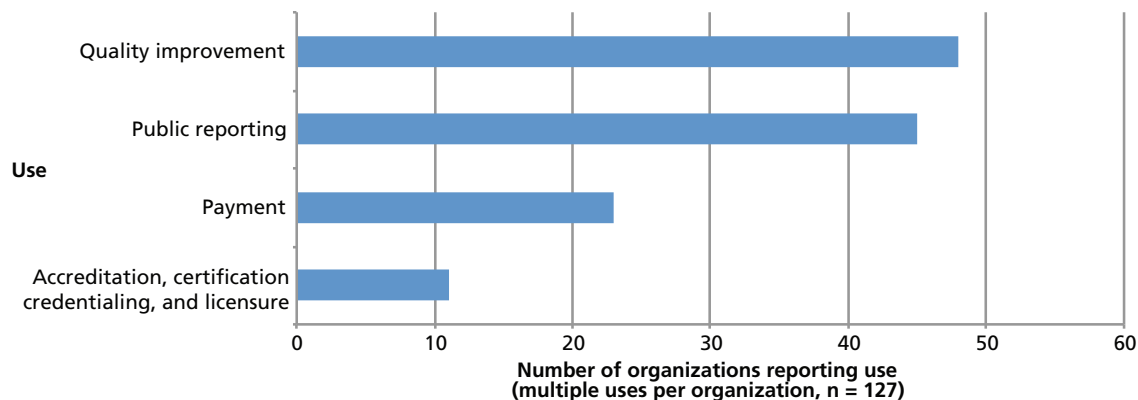
payment uses; 35 percent (16 of 45) of public reporting uses; and 21 percent (ten of 48) of QI uses.
   – In the inpatient setting, the domains most commonly measured are process of care (93 percent), outcomes (73 percent), safety (69 percent), and patient experience (64 percent), while, in the ambulatory setting, the domains most commonly measured are process of care (90 percent), outcomes (68 percent), cost (53 percent) and patient experience (56 percent). (See Table D.4 in Appendix D.)
 • Across 208 uses of measures in various settings of care among the 70 end-user organizations, 69 percent (143 of 208 setting/use pairs) of the uses included a combination of endorsed and nonendorsed measures, 9 percent included only endorsed measures, and 1 percent do not include any endorsed measures (Figure S.2 and Table D.2 in Appendix D). Due to incomplete documentation, we were unable to determine NQF endorsement status for 21 percent (44 of 208) of the setting/use pairs.
 • Claims and administrative data are the most common data source used to construct measures (76 percent [97 of 127] total measure uses reported using administrative data sources), followed by patient survey (56 percent [71 of 127] of uses), self-reported data (38 percent [48 of 127] of uses), and medical records (35 percent [45 of 127] of uses).[2] Each of these data sources was utilized within each of the four broad categories of end use. (See Table D.1 in Appendix D.)

Factors Influencing Measure Use
The following factors affect measure use:

 • A combination of internal and external factors is driving the use of performance measures. For example, federal agencies are responding to legislative requirements related to quality-based payments and public reporting contained in the Patient Protection and Affordable Care Act (Pub. L. 111-148, 2010), and some states face legislative mandates to do public reporting. Some organizations are responding to local health issues, such as obesity, while others indicated they are using measures to operationalize their mission.

---

[2]   We assessed data sources used for each measure use per organization.

Figure S.2
Use of National Quality Forum–Endorsed Measures in Measurement Instances (n = 70 organizations)
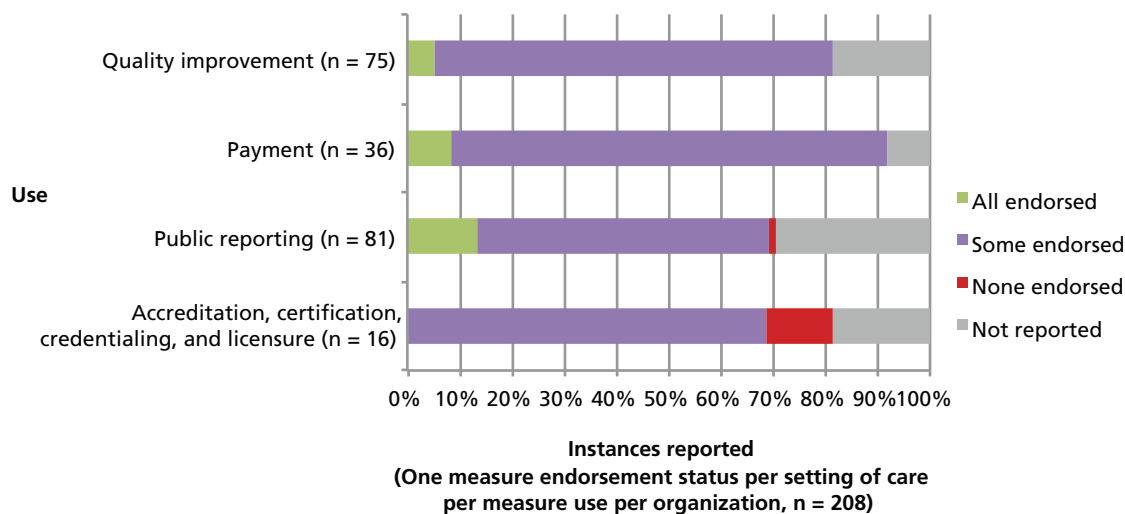
NOTE: The number to the right of each measure use shows the number of discrete organization-setting instances where organizations reported using measures for that purpose.

RAND *TR1148-S.2*

- The single most important factor cited as either facilitating or impeding the use of measures was the availability of data to construct performance measures.
- Factors facilitating the use of measures include a strong data infrastructure; provider trust in the measurement process and the evidence base of the measures; alignment of measures among reporting initiatives to minimize reporting burden; relevance to members, consumers, and providers; and provider training on how to extract the data. Additionally, it was noted that NQF endorsement or widespread use of a measure enhanced provider buy-in.
- In "high-stakes" accountability applications (Safran et al., 2006), such as payment and public reporting, end users reported a preference for using endorsed measures that were evidence-based and validated (among the 30 organizations we interviewed). However, end users indicated a willingness to use nonendorsed measures to address a priority need.
  – The use of nonendorsed measures reflects a perceived need to modify NQF-endorsed measures to adapt to local use (e.g., relax exclusion criteria to increase the number of people included in the measure denominator) or a need to measure an area that is not yet represented in the NQF group of measures (e.g., resource use).
  – Other criteria commonly applied in the selection of measures include relevance to members, consumers, providers, or payers; feasibility of data collection; and scientific acceptability.
- Factors impeding the use of measures include lack of measure prioritization; "measure fatigue" in addressing national, state, and regional requirements (i.e., lack of alignment); the cost of measurement; the lack of timely data; and the challenge of reliably measuring the performance of individual physicians and small physician groups given small denominator sizes for many measures.
- Anticipated changes to the use of measures moving forward include increased capacity for broader and different types of measures with electronic health records (EHRs), reg-

istries and other health information technology (HIT) capabilities; newer types of measures (e.g., longitudinal outcomes, functional status, cost and resource use); new settings (e.g., nursing homes, home health); new populations (e.g., patients who are eligible for both Medicare and Medicaid, referred to as "dual eligible"); and more-widespread use in emerging payment applications (e.g., bundled payments, medical homes, accountable care organizations [ACOs]).

System Changes Resulting from Measure Use

The following system changes resulting from measure use were reported:

- Among those we interviewed, few reported they had performed formal studies to document results, particularly in the realm of return on investment. Several organizations noted that they were just beginning the process to evaluate the impact of their measurement activities.
  – Formal evaluation is hindered by a lack of comparison groups (i.e., control groups, pre/post data), resources to support evaluation, and, in some cases, evaluation "know-how" on the part of the end user.
- Some anecdotal examples were provided of system changes resulting from organizations' use of measures. End users cited improvements in generic prescribing rates, premium savings accrued from implementation of tiered health plan products, performance improvement on selected clinical measures, improvement in patient experience scores, and physician practice changes resulting from QI initiatives.
- Several organizations interviewed reported that performance measures had had little effect on consumer behavior as a result of publicly reporting performance, although they did observe that public reporting has influenced provider behavior.

Advice to Improve the Uptake of Standardized Measures

Our interviews yielded the following recommendations for improving the uptake of standardized measures:

- A range of measure gaps and opportunity areas for new measures were identified, including coordination of care, outcomes, patient safety (inpatient and ambulatory), measures of longitudinal change in outcomes, cost and resource use, population health, nursing-sensitive care, access and affordability, measures that consider a patient's risk profile, composite measures, and measures that address specific conditions (e.g., maternity, mental health, end-of-life care) and settings and populations (e.g., home health, dually eligible individuals).
- There were recommendations for adding new measures that address specialty care; however, some respondents commented that resources should not be used to develop unique measures for every specialty. Consideration should be given to developing crosscutting measures that address the care delivered by multiple clinical specialties.
- The expression of need for more measures was counterbalanced by comments regarding the need for better alignment and prioritization of existing measures to help in selection and use of measures. One interviewee commented, "It is less a demand for more measures and more a problem that the extant measures are very distracting and not aligned." Interviewees stated a need to ease their selection of measures by "weeding out" or rationalizing

endorsed measures or selecting the "best in class" when there are very similar endorsed measures, as well as guidance on choosing the most important and effective measures to drive change.

- There were specific recommendations from end users related to opportunities for NQF process improvements:
  - End users expressed a desire for improvements to the NQF website to enhance the search capability, obtain full measure specifications, and access more information about where and how measures are being used.
  - End users also suggested that NQF work to advance the conversation about the appropriate use of composite measures, provide a clearinghouse for QI measures, provide outreach and education to stakeholders in their communities (particularly physicians and other health-care providers), publicize upcoming calls for measures and the review calendar so that organizations could coordinate their review processes with NQF's, and provide more advanced notice of upcoming calls for measures so that developers have more time to test measures prior to submission.

Future Opportunities to Support the Use of Measures

As the measurement development, endorsement, and application communities reflect on the results of this study, we note that several issues emerged that represent opportunities and potential areas of focus moving forward to help facilitate the use of measures:

- *Establish priorities for where end users should focus their attention and resources.* Although NQF has a large number of endorsed measures, the sheer number is daunting to end users, and many measures are viewed as irrelevant or low impact. Given the limited resources and attention that end users can devote to measurement, they need guidance on defining a more limited set of high-value, high-leverage measures to which they should devote resources that will lead to real improvements in outcomes or efficiency (fewer resources spent without reductions in quality). Outcome measures were seen as high-value measures that should be prioritized.
- *Align measures.* Among those who are the focus of measurement (i.e., providers), there is an important need to help reduce the burden of reporting to a wide array of entities by ensuring that the measures used by the entities are aligned. Currently, reporting efforts among the various stakeholders operate independently, are frequently not aligned, and lead to increased data collection, analysis, and reporting burdens for providers. Establishing measure priorities can create an opportunity for better alignment among the various parties that impose reporting requirements. Cataloging the specific types of measures being used for each end use could help inform efforts to align measures.
- *Develop new measures for new measure uses.* The measure-use landscape continues to evolve, particularly in response to health system reform efforts. End users are preparing for new measure-use opportunities and the need for different types of measures, particularly in the payment use area (e.g., ACOs, value-based purchasing [VBP] applications, episode-of-care and bundled payment models). Additionally, there is a desire for cross-cutting measures that apply to all providers both to broaden the group of providers that can be assessed and to minimize the number of measures that have to be constructed. Additionally, end users expressed strong interest in new measures that would track the

functioning of the patient and how functioning and other key outcome markers change longitudinally (i.e., improvement measures).

- *Build support for the use of measures.* The criteria that are most salient to end users in choosing measures for use are (1) the measure is relevant to providers who must act on them, as well as consumers and payers; (2) the data needed to construct the measure are feasible to collect; and (3) the measure has a scientific evidence base to ensure its acceptability to providers.

- *Measure construction requirements in the evolving data landscape.* At present, end users are relying heavily on claims and administrative data and patient surveys and are not yet able to systematically and efficiently capture information that will enable the construction of intermediate and longitudinal outcomes. At a minimum, a limited, prioritized set of measures should be the focus of front-end planning to influence the data architecture (e.g., data fields and not free text) of EHRs and other HIT to support the construction of measures. The development of better data sources, such as what is envisioned through HIT sources, has the potential to enable better measurement and use of measures.

- *Conduct a systematic review of the literature to fully catalog measure use.* The scope of this project did not permit a full review of the literature to examine uses of measures for a variety of purposes and to explore the secondary use of measures (e.g., by providers who are exposed to performance results in public report cards, consumers who are shown performance report cards). A thorough review of the literature would extend the work in this study and provide for a deeper understanding of uses and issues related to use of measures.

- *Formally assess the system-change results from the use of measures.* Although end users cite anecdotal examples of the benefits that have accrued from use of performance measures, it is unclear how many of these benefits have been formally evaluated and documented. This underscores the need for a future study that would systematically attempt to quantify the impacts that have resulted from measure use, given the substantial investment that has been made to use measures to drive system change.

- *Create support tools to help end users.* End users want access to full measure specifications, more information about where and how measures are being used, and an ability to more easily search the measures contained in the NQF-endorsed set of measures. Improved search functions and new tools could facilitate end users' ability to use measures. End users also see value in increased outreach and education to stakeholders in their communities (particularly physicians and other health-care providers) about the validity of the measures and why measurement is important as a means of better engaging them in performance improvement.

Acknowledgments

Abbreviations

ACO	accountable care organization
CAHPS	Consumer Assessment of Healthcare Providers and Systems
CMS	Centers for Medicare and Medicaid Services
CPOE	computerized physician order entry
CVE	Chartered Value Exchange
EAP	Evaluation Advisory Panel
EHR	electronic health record
ER	emergency room
HCAHPS	Hospital Consumer Assessment of Healthcare Providers and Systems
HEDIS	Healthcare Effectiveness Data and Information Set
HHS	U.S. Department of Health and Human Services
HIT	health information technology
MOC	maintenance of certification
NCQA	National Committee for Quality Assurance
NQF	National Quality Forum
P4P	pay for performance
QI	quality improvement
ROI	return on investment
VBP	value-based purchasing

Introduction

Purpose

Over the past two decades, there has been substantial growth in the number of entities using health-care performance measures for a host of purposes. Quality and efficiency measures are now embedded throughout the U.S. health-care system, as demonstrated by their widespread use for accountability; accreditation, certification, credentialing, and licensure; public reporting; pay for reporting (P4R); pay for performance (P4P) and performance-based contracting; tiering and construction of narrow provider networks; quality improvement (QI); and public recognition. Performance measures are being used by the following:

- providers and integrated health systems for QI and for use internally in incentive-based performance schemes with their contracting providers, to strengthen their ability to succeed in a performance-based contracting environment
- accrediting organizations for ensuring high standards of care by health plans, hospitals, and other types of facilities
- public and private purchasers of care (and their health plan agents) for public accountability, consumer decisionmaking, performance-based contracting, construction of new payment methods that incentivize high-performing providers, and structuring other incentives to drive patients toward higher-value providers (e.g., tiering, narrow provider networks)
- community collaboratives (e.g., Chartered Value Exchanges [CVEs], Aligning Forces for Quality communities) for public accountability, enhancing awareness among consumers about key health decisions and choices among providers, and QI through peer pressure and public recognition
- state agencies to assist patients and consumers with accessing comparative information on health-care providers and to promote public accountability
- professional boards as part of new certification and maintenance-of-certification (MOC) programs for individual providers
- research organizations and foundations to monitor improvements in care delivery and the extent to which unintended consequences are occurring and to construct report cards that compare provider performance for use by patients and their families.

The underlying feature of all of these applications is a focus on using measurement and reporting to drive fundamental delivery-system changes across all levels of the health system.

Although there has been significant growth in the use of performance measures, there has been no systematic evaluation to catalog the various ways in which measures are being

used (Madison, 2009; National Committee for Quality Assurance, 2010), the factors influencing the uptake of measures, whether changes have occurred in the health system changes as a result of the use of the performance measures, and whether there are missed opportunities for enhancing end users' ability to use performance measures to achieve their desired objectives. The U.S. Department of Health and Human Services (HHS) requested that the National Quality Forum (NQF) work with an independent evaluator to examine the uses of performance measures. The purpose of the project was to gather information directly from end users of measures to inform the efforts of the federal government, NQF, measure developers, and others engaged in the quality-measurement enterprise, all of which collectively contribute to the portfolio of measures that can be used by a wide array of end users.[1]

Study Objectives and Evaluation Framework

The objectives of the study were to do the following:

- Evaluate the current state of the use of NQF-endorsed performance measures in the field.
- Determine ways in which performance measure are being used:
  - accreditation, certification, credentialing, and licensure
  - public reporting
  - QI
  - payment, network selection (e.g., P4P incentives, tiering)
  - other.
- Begin to explore the impact of performance-measure use on system change as groundwork for a future study to assess the impact of the use of measures.
- Identify opportunities to support the use of measures.

To frame the evaluation, NQF staff developed a logic model depicting the expected outcomes of using standardized health-care performance-measurement systems. The logic model builds on theoretical constructs from the ground-laying work of the Strategic Framework Board (McGlynn, 2003) and subsequent Institute of Medicine studies (Institute of Medicine, 2001, 2006). At the start of the project, RAND and NQF, in consultation with the Measure Use Evaluation Advisory Panel (EAP) (see Appendix A) worked to finalize the model, which is shown in Figure 1.1. Notionally, the model illustrates a process by which the use of standardized performance measures ideally should (1) promote systems of accountability, (2) facilitate continuous QI processes, and (3) improve systems that support population health.

The scope of work in this study focused on the area of the model labeled "measure uses" and "influencing factors." Potential factors that might influence the adoption of performance measures include legislative mandates, the underlying data infrastructure to support measure construction, provider buy-in to measures for a given purpose, professional commitment to QI, and market forces. The area to the right of "measure uses" describes broad system changes that represent the desired intermediate outcomes from the use of performance measures. The system changes that occur from use of measure fell outside the scope of this project; however,

---

[1]   As required under the Medicare Improvements for Patients and Providers Act of 2008.

Figure 1.1
Final Performance Measure-Use Logic Model
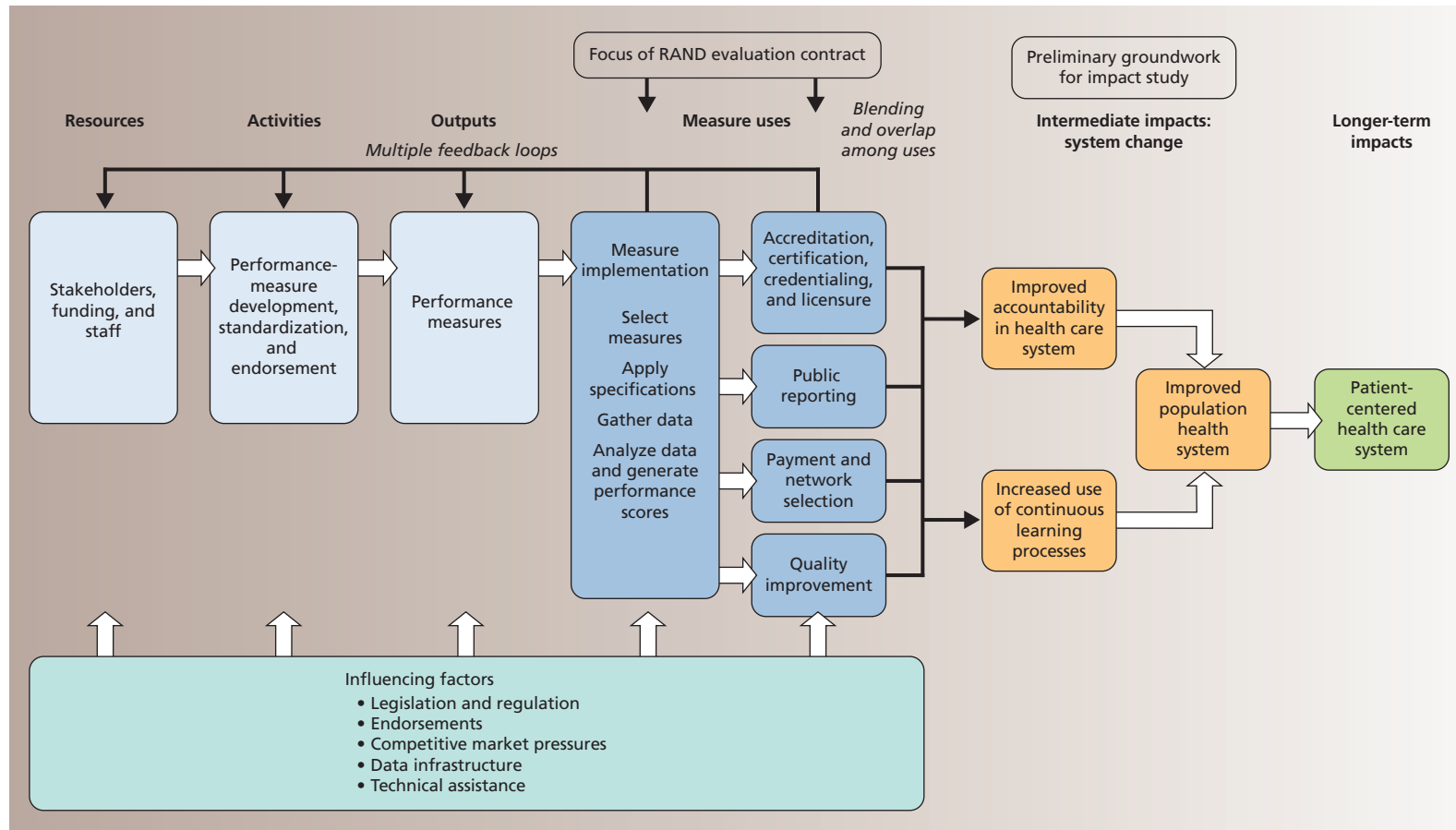
we did briefly explore this topic with end users as preliminary groundwork for a future study that would more comprehensively assess the impact from measure use.

Study Methods

In this study, we focused on understanding measure use related to four broad purposes or end uses:

- accreditation, certification, credentialing, and licensure
- public reporting
- QI
- payment and network selection.

If organizations stated they were using measures in other ways, we also captured the other uses mentioned. Although there was strong interest in learning about the use of NQF-endorsed measures, RAND also explored the use of nonendorsed measures for the different end uses and the rationale for using nonendorsed measures.

Evaluation Topics

The study addressed the following four overarching questions, with subtopics listed below each question:

- **How are performance measures being used in practice?**
  - What types of measures are being used?
  - What factors affect use of NQF- versus non–NQF-endorsed measures?
  - To what extent are organizations using measures that can be constructed from health information technology (HIT) or clinically enriched administrative data sources versus measures that are constructed using claims and encounter data?
- **What factors are influencing measure use?**
  - What are facilitating factors?
  - What are the barriers to measure use?
  - What role do external factors play in influencing the use of measures?
  - What roles do the data infrastructure and HIT play in use of measures?
- **What types of system changes have occurred as a result of measure use?**
  - What benefits or results have been accrued (e.g., effects on quality, costs of care, how care is delivered)?
  - What adverse effects have been associated with measure use?
- **What improvement areas could facilitate the use of standardized endorsed performance measures by various end users?**

– What are the challenges or missed opportunities in which the use of performance measurement is not being fully realized?
– What can be done to create new opportunities for performance-measurement use, and how can different systems be reconciled?

Defining End User

We focused our review on the primary end user of performance measures. The primary end user is defined as any organization that is directly engaged in the implementation of health-care quality or efficiency measures to assess the performance of providers of health care (i.e., hospitals, physicians, nursing homes, health plans). The primary end user is the organization that gathers the data; gains access to data for use in constructing measures or, in some cases, uses preexisting measures; uses data to construct measures; computes the performance score of providers; and makes use of the resulting performance information. The primary end user makes performance information available with the expectation that others (i.e., providers, consumers) will use the performance measures to drive system changes, make behavior changes, or inform decisions. The "others" are secondary end users of the information. An example of the secondary use of performance measures would include a medical group changing its behavior in response to receiving performance scores from a community collaborative or consumers using comparative performance information on health plans to select a health plan.

We excluded from our review secondary end users, which we defined as organizations that make use of existing performance-measure results provided or reported by a primary end user (e.g., a medical group acting on a report card produced by a community collaborative), that advocate for use of performance measures, develop measures for use by others but do not actually use the measure (e.g., the American Medical Association's Physician Consortium for Performance Improvement), or provide technical assistance or consulting services to help health-care providers improve their performance on measures. The scope of this contract did not permit the exploration of secondary uses, although this would be an important area to address in a future impact study.

Defining Use

An organization was deemed as using performance measures (coded in our data as "1 = yes") if there was evidence that the organization was using performance measures for one of the four key purposes: QI, payment, public reporting, or accreditation, certification, credentialing, or licensure. This information was derived from key-informant interviews and documentation available from the organization. Each organization could be coded for a maximum of four different uses. Within a use, an organization could be coded as using measures in multiple settings (use/setting pairs, such as P4P in a hospital setting or P4P in an ambulatory setting).

We did not attempt to evaluate differences among organizations in the level or degree to which they were using measures for a similar purpose; this was beyond the scope of this project. As such, when interpreting the findings from this study, the same reported uses are not necessarily equivalent in their level of penetration, number of measures used, or other factors that might be used to assess differences among organizations within a given end use.

Data Sources

The study design consisted of two data-collection methods: (1) interviews with end-user key informants (n = 30 end-user organizations) and (2) review of publicly available documents and materials from websites (n = 70 end-user organizations, of which 30 participated in the key-informant interviews).

Sample Frame

To construct a sample frame (i.e., the universe from which the sample of organizations was drawn), we developed a list of organizations that were involved in quality measurement and were potential end users of performance measures. We compiled information from multiple sources to create a comprehensive list of end users. We started with the NQF member list and augmented this by drawing from

- the expert knowledge of members of the research team who previously had conducted research on the use of performance measures and have worked with organizations using performance measures
- input from the project's EAP, which included representatives from various end-user categories, such as purchasers, consumers, providers, and accrediting bodies
- information from the Leapfrog Group's Incentive and Reward Compendium
- background materials from other studies of measure uses provided by NQF (i.e., a Mathematica catalog of public reporting programs and measures and the study of state Medicaid program efforts) (O'Neil, Schurrer, and Simon, 2010; Highsmith, Mahadevan, and Lind, 2011)
- the National Academy for State Health Policy, to identify state-based initiatives
- discussions with the America's Health Insurance Plans and the Blue Cross and Blue Shield Association
- organizations suggested by NQF and HHS staff
- a focused Google search around the four end uses identified in the logic model (i.e., accreditation, consumer information and public reporting, payment, QI). We combined these end uses to create a variety of search strings to identify specific organizations that are using measures for specific end uses. An example of a search is one that combined "quality measures" and "accreditation." These terms also were combined with the organization type and health-care setting to further focus our searches.

From these different sources, we generated a list of approximately 580 organizations that, in some manner, were engaged in performance measurement.

We created 11 end-user categories to reflect different types of organizations that use performance measures. The 11 categories were

- community collaboratives (includes CVEs and Aligning Forces communities)
- purchasers and purchaser coalitions
- consumer groups
- health plans (payers)

- state government agencies
- federal government agencies and departments
- accreditation, certification, credentialing, and licensure bodies
- provider specialty organizations
- health-care improvement organizations and QI organizations
- providers (hospitals, health systems, physicians)
- other (e.g., foundations, media).

Each of the 580 organizations was categorized into one of the 11 organization types. For each organization in our master list, we initially denoted the purposes for which it is using performance measures or engaged in performance-measurement activity to the extent that this information was available or determinable based on a cursory review of public documents. Organizations could use measures for up to four different measure uses (QI, public reporting, payment, and accreditation, certification, credentialing, and licensure). The missions of the end users varied across organizational categories and sometimes within categories. We applied our end-user definition and excluded from this larger list organizations that were deemed secondary users of performance information.

Sample Selection Criteria and Method
Because the goal of the study was to gain an understanding of the wide range of end uses of performance measures and the types of measures being used within each type of end use, we used a purposive sampling approach to balance both different types of end uses and different types of end-user categories. In selecting the sample, we considered the following:

- ensuring broad representation across the end-user category types (e.g., consumers, purchasers, states). We wanted to ensure that each end-user category was represented for at least one measure use. In some cases, a single organization was selected for more than one measure use (e.g., the National Committee for Quality Assurance [NCQA] uses measures for public reporting and accreditation).
- ensuring broad representation of measure use within different types of health-care settings and for different populations
- allocating the sample in roughly equal amounts across the four broad types of end uses to the extent that there were sufficient numbers of end users available for a given end use.

We purposefully selected 70 primary end-user organizations, starting by identifying 30 organizations for the key-informant interviews. These 70 organizations were selected in a manner that would ensure representation across the 11 end-user categories because these different end users deploy measures for different purposes given their organizational missions. The missions of the end users included organizations focused on improving quality, creating information to help consumers select providers, working to assess the competencies of providers and health plans to deliver high-quality care, structuring incentives to drive improvements in care, and promoting competition on price and quality.

The 30 key-informant organizations were also included in the document and website review component of the study. Having selected the initial 30 organizations, we then drew an additional 40 organizations. The 40 organizations underwent only the document and website review. The list of 70 organizations included in the sample is contained in Appendix B.

It should be noted that the sample was not designed to be representative of either the various end uses or types of end users; rather, the sampling approach was designed to allow the study team to gather information for the different types of end users and end uses. Drawing a random sample would prove difficult. As such, we worked to ensure that we populated the cells of a table that contained end-user categories as the rows and end uses as the column. Not all end-user categories were represented for each measure use because, in some instances, members of that end-user category do not engage in that activity. The purpose of this project was not to assess the prevalence of use of measures by either end-user category or end use, which a random sample would have enabled. Furthermore, had we drawn a random sample, we would have potentially omitted some end users or end uses due to their low prevalence in relation to the total population of end users or end uses (e.g., foundations or media).

Table 2.1 displays the number of organizations from each end-user category represented in the document and website review and key-informant interview samples.

Interviews

We conducted semistructured phone interview with 30 key-informant organizations. We interviewed the executive director, president, or chief executive officer or the lead person responsible for quality-measurement activities within the organization. In the event that responsibilities were split across multiple people, we either conducted multiple interviews for a single organization or involved multiple people in a single interview. Interviews were approximately one hour in duration and were tailored to the type of organization. The types of topics covered in the interviews included

Table 2.1
Representation of Organization Categories in Review Samples

Organization Category	Document Review Sample	Key-Informant Interview Sample
Accreditation, certification, credentialing, and licensure bodies	5	3
Community collaboratives	10	3
Consumer groups	2	2
Federal government	4	2
Health-care improvement organizations	4	2
Health plans (payers)	11	4
Provider specialty organizations	6	3
Providers	8	4
Purchasers and purchaser coalitions	10	3
State government	8	2
Other (e.g., media, foundations)	2	2
Total	70	30

- ways in which the organization used performance measures and the types of measures used
- factors driving the use of performance measures
- the process used to select measures and factors influencing the adoption of measures
- ways in which performance measures were integrated into the organization's work and any modifications to existing measure specifications
- factors that facilitate or impede the organization's ability to readily use performance measures
- anticipated changes to measurement activities in coming years
- system changes resulting from use of performance measures, both positive and negative
- areas in which the use of performance measures could be expanded
- how the organization successfully increased the use of performance measures
- ways to improve the uptake of standardized performance measures.

Interviews were recorded and transcribed. Key themes were coded from the transcripts.

Website and Document Reviews

We reviewed documentation and the websites for all 70 organizations in the sample. In the event we were unable to locate information on the organization's website or if the information found did not appear to be complete based on the team's knowledge of the organization's performance-measurement activities, we identified and contacted the person within the organization responsible for performance measurement and requested available documentation about its activities. We received additional documents for many of the organizations that we reviewed.

We systematically coded information on how measures are being used. Each organization in our sample could be coded as engaging in any or all of the four measure uses of interest:

- An organization was coded as a "1 = yes" for each use it reported for a maximum of four different types of uses per organization.
- Within each measure use, we recorded the type of provider whose performance was being measured (e.g., hospital, physician group, individual physician) as the unit of measurement, the setting in which measured care was delivered (e.g., inpatients, ambulatory, nursing home), and the types of data being used (e.g., claim data, patient surveys, medical record data).
- It was common for each measure use by an organization to focus measurement activities on multiple types of health-care providers, include multiple settings of care, and use multiple types of data.
    - Within a single use (such as payment), an organization could be coded as having multiple instances of use. For example, an organization could be using measures for P4P in the hospital setting and in the ambulatory setting (i.e., setting/use pairs). In this situation, the organization would be coded as having two instances within the payment end use.
    - Furthermore, these two settings have two different types of providers whose performance is being measured—hospitals and physician organizations—so again, the orga-

nizational entity would report multiple unit of accountability/use instances as to the unit to which performance is being attributed.

– Last, the organization might rely on a mix of patient surveys, medical record review, and administrative claims to construct measures (i.e., data source/use instances).

– For each measure use in each setting, we drilled down to assess the types of measure domains being included in measurement efforts (e.g., process, outcomes, patient experience) and the extent to which NQF-endorsed measures were being used (e.g., only endorsed measures, some endorsed measures, no endorsed measures).

  • In some cases, organizations did not report the NQF endorsement of their measures (21 percent instances of measure setting/use pairs) and did not provide enough detail for us to determine the source of the measures. This ranged from 8 percent of payment efforts to almost 30 percent of public reporting activities).

Appendix C illustrates these various combinations. The 70 organizations in the sample reported 127 distinct measure uses (i.e., instances), and, within the 127 uses, there were 236 distinct units of measurement/use instances, 208 setting/use measure instances, and 261 data source/use instances. Appendix D contains tabular results summarizing additional findings.

We did not establish any minimum criteria that had to be met for an organization to be coded as engaging in a measure use other than the organization's representative stating to us that it did so or evidence provided through documentation. We also did not develop criteria to assess the degree or level of use for a given purpose to determine whether organizations were using measures in equivalent ways. Thus, there is likely substantial variation across the organizations in the extent to which and sophistication with which they are using performance measures for each use.

Study Limitations

There are several limitations of the study that should be considered when interpreting the results. First, the purposive sampling approach we used was selected to help illustrate the range of uses of measures by different types of end users. Although the sample is diverse in its composition, it is unclear to what extent the findings from this sample of organizations would generalize more broadly to the full universe of end users.

Because of the small number of organizations of any single type included in the interviews, our ability to compare and contrast the findings across organization types was limited. Also due to the small number of interviews, our results are not generalizable to the broader community of organizations using performance measures. The participants in the study are likely among the more experienced users.

Another limitation of this study is that the findings presented might underestimate the total number of uses and types of uses by the organizations selected for inclusion in the study. This could occur if the information that was available through public documents was not a complete representation of an organization's full set of performance measures and uses. In some cases, organizations view this type of information as proprietary or have not fully posted all information on their websites or in public documents. To the extent that this occurred, our estimates would undercount the true number and types of uses in the organizations in the sample. To mitigate this potential problem, we performed follow-up via telephone outreach to

organizations to request additional materials when the study team did not see documentation or when the team felt that documentation was incomplete.

In the context of this study, we did not attempt to classify the different levels or degrees of use of measures by end users for the same purpose. For any single type of use, there is undoubtedly variation across end users, such as in the robustness of measures used, penetration level among the population of providers that could be measured, and the number and type of measures used for a particular purpose. This differentiation of use within an end use was beyond the scope of this project but could represent an important area of exploration for a future project.

Study Findings

In summarizing the findings that address the first research question, "How are measures of performance being used in practice?" we draw information gleaned from both the key-informant interviews and the document review. For research questions 2–4, our findings are based on what we learned from the key-informant interviews because the document review did not generate information to address these questions.

Question 1: How Are Performance Measures Being Used in Practice?

In conducting our review of publicly available documentation on the use of performance measures, we cataloged uses by 11 different types of end-user organization categories. We identified 127 discrete uses among 70 organizations, with more than half reporting multiple uses. Forty out of the 70 organizations reported two or more uses. Only one organization reported the use of measures for all four purposes. Health plans, providers, and federal agencies used measures for a wider variety of purposes, whereas consumer groups and health-care improvement organizations focused on just one use.

Table 3.1 shows the organization types and number of organizations reporting one, two, three, or four different end uses.

Among the organizations in this sample, QI (n = 48) and public reporting (n = 45) were the most frequent uses identified, followed by payment (n = 23) and accreditation, certification, credentialing, and licensure (n = 11) (Figure 3.1). Several organizations that we interviewed noted they use measures for other purposes, including monitoring the compliance of health plans and evaluating the impact of interventions. We did not see examples of other uses of measures in our review of public documents; however, it is possible that other uses were present and the documentation that was available to our review team did not contain that information.

Each measure use by an organization could

- include multiple settings of care
- use multiple types of data
- focus measurement efforts on multiple types of health-care providers.

For example, an organization could have public reporting programs that focus on inpatient and ambulatory care. These two different settings represent two different measure-use instances. Furthermore, these two settings have two different units of measurement—hospitals and physician organizations—so again, the organizational end user would report multiple

Table 3.1
Number of Measure Uses, by Organization Type (n = 70 organizations)

Type of Organization	1	2	3	4
All organizations	30 (43%)	24 (34%)	15 (21%)	1 (1%)
Accreditation, certification, credentialing, and licensure bodies	2	1	2	0
Community collaboratives	4	5	1	0
Consumer groups	2	0	0	0
Federal government	0	2	2	0
Health-care improvement organizations	4	0	0	0
Health plans (payers)	3	2	5	1
Provider specialty organizations	1	4	1	0
Providers	1	5	2	0
Purchasers and purchaser coalitions	7	2	1	0
State government	5	2	1	0
Other	1	1	0	0

Figure 3.1
Frequency of Measurement Uses in Document Review Sample (n = 70 organizations)
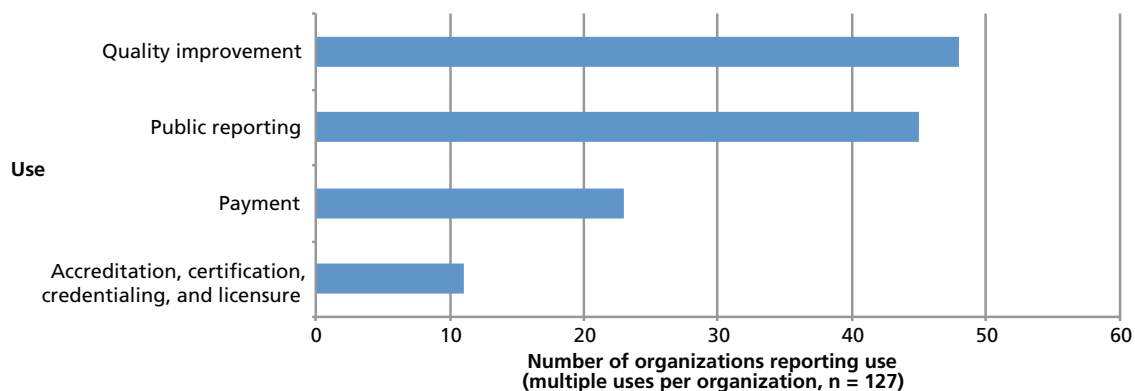
NOTE: An organization can report that it was using measures in one or more of these end-use categories, up to a maximum of four different end uses per organization.
RAND *TR1148-3.1*

instances for the unit of measurement being evaluated. Last, the organization might rely on a mix of patient surveys, medical record review, and administrative claims to construct measures. The frequencies reported in the remainder of this section show these combinations. Appendix C illustrates these various combinations. The 70 organizations in the sample reported 127 distinct measure uses (i.e., instances), and, within the 127 uses, there were 236 distinct units of accountability. Appendix D contains tabular results summarizing additional findings.

Figures 3.2 through 3.5 present the frequency with which the different types of organizations used measures for each of the four measure end uses. The number to the right of each organization-type description shows how many organizations of that type were included in the

Figure 3.2
Use of Measures for Quality Improvement, by Organization Type (n = 70 organizations)

Organization type

■ Reports use   ■ No report of use

Organization type	
Accreditation, certification, licensing, and credentialing (n = 5)	
Community collaboratives (n = 10)	
Consumer groups (n = 2)	
Federal government (n = 4)	
Health care improvement orgs (n = 4)	
Health plans (n = 11)	
Provider specialty organizations (n = 6)	
Providers (n = 8)	
Purchasers and purchaser coalitions (n = 10)	
State government (n = 8)	
Other (n = 2)	

Number of organizations

RAND *TR1148-3.2*

Figure 3.3
Use of Measures for Public Reporting, by Organization Type (n = 70 organizations)

Organization type

■ Reports use   ■ No report of use

Organization type	
Accreditation, certification, licensing, and credentialing (n = 5)	
Community collaboratives (n = 10)	
Consumer groups (n = 2)	
Federal government (n = 4)	
Health care improvement orgs (n = 4)	
Health plans (n = 11)	
Provider specialty organizations (n = 6)	
Providers (n = 8)	
Purchasers and purchaser coalitions (n = 10)	
State government (n = 8)	
Other (n = 2)	

Number of organizations
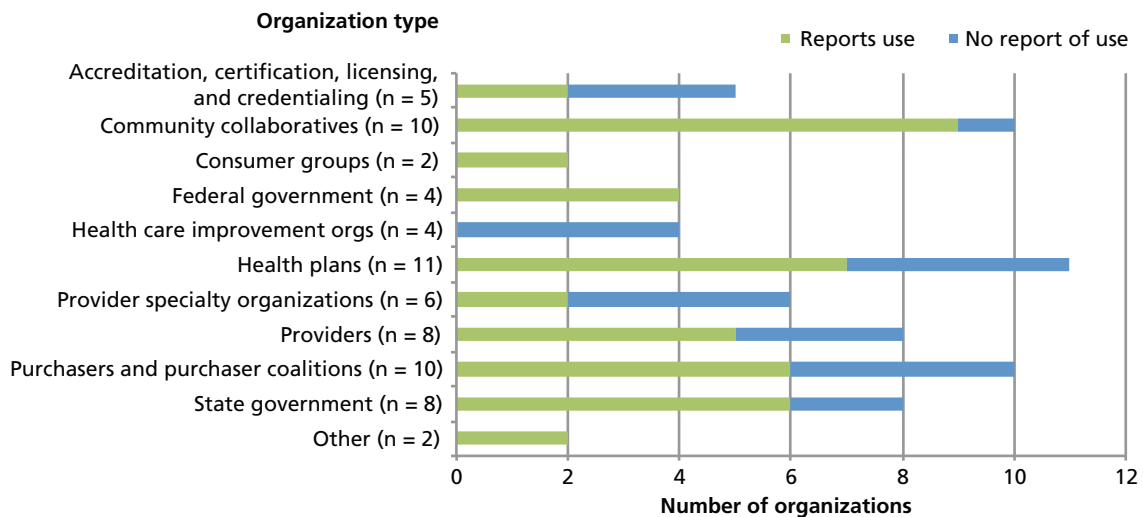
RAND *TR1148-3.3*

sample. The bars show the number of organizations within that organizational category that reported use of measures for that particular end use.

Figure 3.2 illustrates the frequency with which different types of organizations used measures for QI. All end-user organization types with the exception of consumer groups used measures for QI.

Figure 3.3 displays the use of measures for public reporting by organization type. Organizations from each organization type except health-care improvement organizations used measures for public reporting. In the sample we drew, consumer groups, federal and state govern-

Figure 3.4
Use of Measures for Payment, by Organization Type (n = 70 organizations)
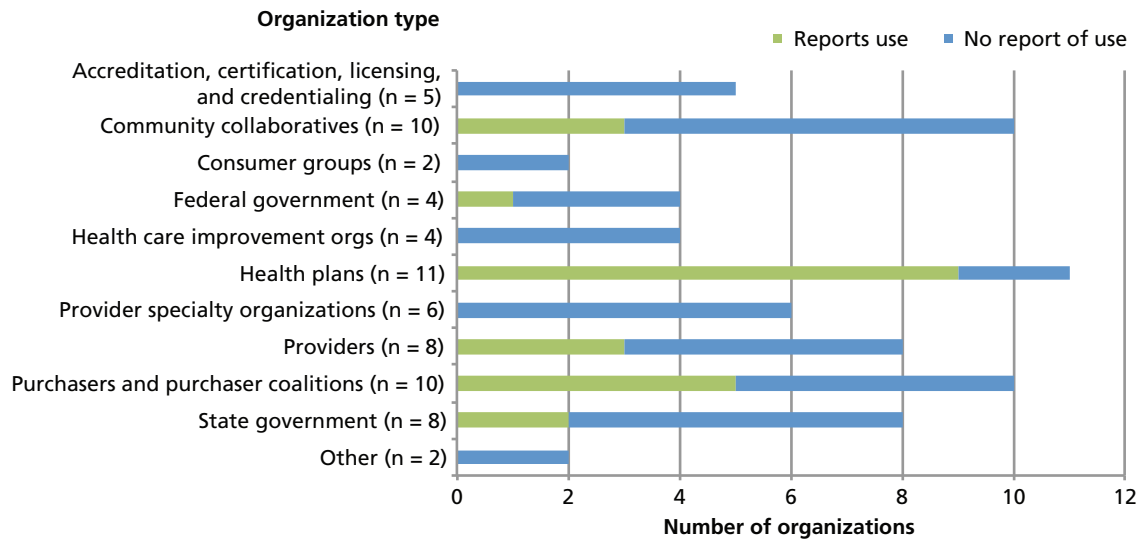
Figure 3.5
Use of Measures for Accreditation, Certification, Credentialing, and Licensure, by Organization Type (n = 70 organizations)
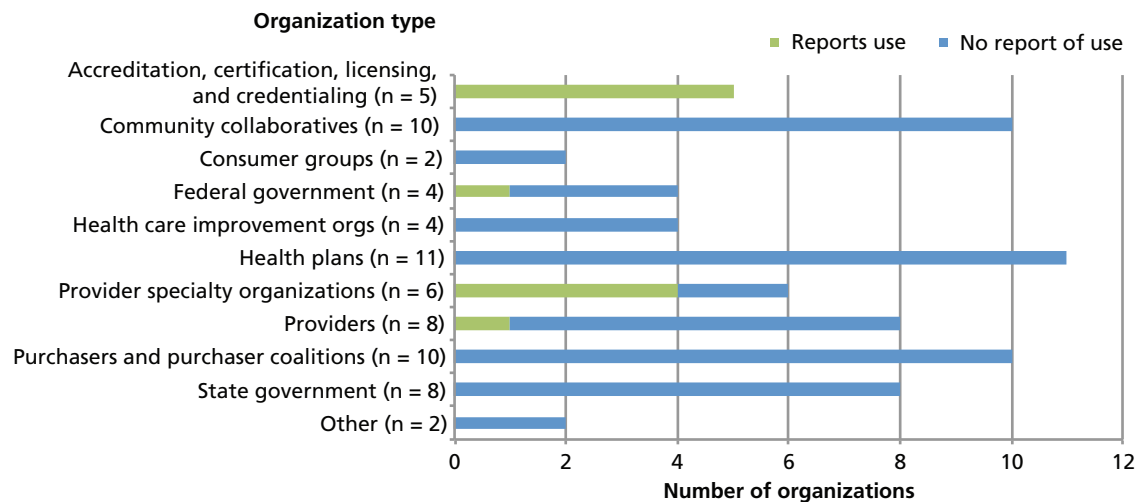
ment agencies, community collaboratives, and "other" (e.g., media and foundation) were most likely to use measures for public reporting.

Figure 3.4 presents the frequency with which organizations in the sample included measures into their payment applications. Although more than three-quarters of health plans and half of the purchaser and purchaser coalition organizations in the sample incorporated the use of measures into their payment approaches, only a minority of other organization types did so.

Consumer groups; accreditation, certification, credentialing, and licensure; provider specialty organizations; and "other" did not report using measures for payment purposes.

Figure 3.5 shows the types of organizations that use measures for accreditation, certification, credentialing, or licensure. All of the accreditation, certification, credentialing, and licensure bodies, as well as two-thirds of the provider specialty organizations, use measures for accreditation, certification, credentialing, or licensure. There were a few provider organizations and federal agencies that reported use of measures for this purpose; otherwise, there was no reported use of measures for this purpose by other organization types.

Settings Included in Measurement Efforts

An organization's reported measure uses could each involve multiple settings of care (e.g., public reporting in the ambulatory setting and the hospital setting). Across the 70 organizations in our sample, there were a total of 208 setting/use pairs identified. When the total is averaged across the 70 end users, the average number of settings measured per organizational entity was 1.8 settings. Within the sample, the ambulatory and inpatient settings were the most frequent reported settings as the focus of performance-measure use (Figure 3.6).

Types of Measures Used in Measurement Efforts

We recorded the measure domains and use of NQF-endorsed measures at the level of the setting (e.g., inpatient) within the measure use (e.g., public reporting). This approach was used to capture different components of measure uses because each measurement effort might include a different mix of measure domains. For example, an inpatient performance-measurement program might include measures from the process, safety, and patient-experience domains of care, while an outpatient performance-measurement program might include measures from the process, outcome, and structure domains. We coded each combination as an instance of performance-measure use.

Figure 3.6
Settings of Care Included in Measurement Efforts (n = 70 organizations)
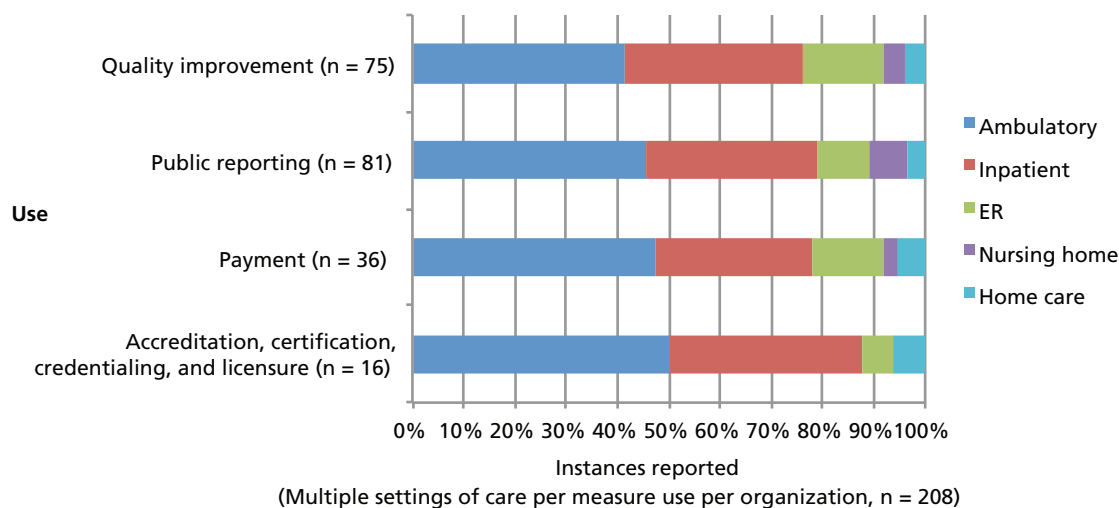
NOTE: The number to the right of each measure use shows the total number of instances of measure use being reported for that particular use. The bars show the percentage of instances involving each setting for each specific measure use. ER = emergency room.

RAND TR1148-3.6

Most uses of measures involved a mix of measure types (domains of care) and a mix of endorsed and nonendorsed measures (70 percent of all setting/use instances) (Figure 3.7 and Table D.2 in Appendix D). Among those end uses for which adequate information was available to make a determination, only NQF-endorsed measures were used in 9 percent of instances of measure use. Almost all of the instances of measure use included an NQF-endorsed measure; only 1 percent of all measure-use instances had no NQF-endorsed measures. However, an important caveat is that, in some cases, organizations did not report the NQF endorsement of their measures (21 percent of setting/use instances of measure uses) and did not provide enough detail for us to determine the source of the measures.

Measures from seven domains (structure, process, outcome, access, safety, cost, and patient experience) were used for each measure use (Figure 3.8 and Table D.4 in Appendix D). With 743 individual mentions of domains within 208 settings within measure uses in the sample, organizations reported an average of 3.6 domains measured per setting. Process measures were most common and were used in 89 percent (186 of 208) of measurement instances identified (as defined by setting/measure use combinations). Outcome and patient experience measures were reported used in more than half of all setting/measure use combinations (136 of 208 and 118 of 208, respectively). Safety measures were used in 48 percent of all setting/measure use instances (100 of 208). Access measures were the least frequently used and were included in 18 percent (36 of 208) of setting/measure use instances.

For the key-informant interviews, we were able to drill down to understand the specific measures being used within domain. The most commonly used measures for assessing performance in the inpatient setting include those reported on Hospital Compare (HHS, undated), the Patient Safety Indicators (Agency for Healthcare Research and Quality, undated), Leapfrog patient-safety measures, surgical-site and central-line infections, the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) survey (HCAHPS, undated), and various registry measures. Measures being used to assess performance in the ambula-

Figure 3.7
Use of National Quality Forum–Endorsed Measures in Measurement Instances (n = 70 organizations)

Instances reported
(One measure endorsement status per setting of care
per measure use per organization, n = 208)

NOTE: The number to the right of each measure use shows the number of discrete organization-setting instances where organizations reported using measures for that purpose.

RAND *TR1148-3.7*

Figure 3.8
Use of Measure Domains in Measurement Instances (n = 70 organizations)
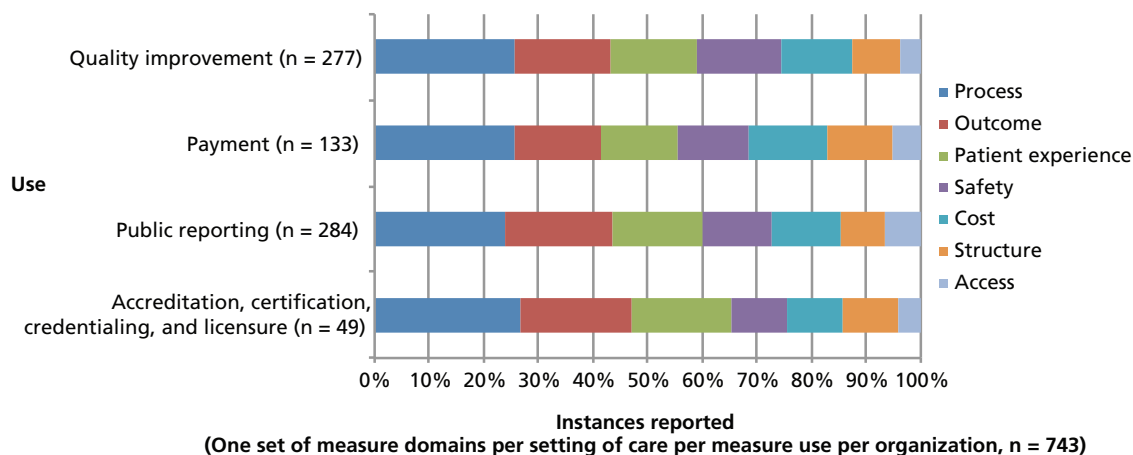
Instances reported
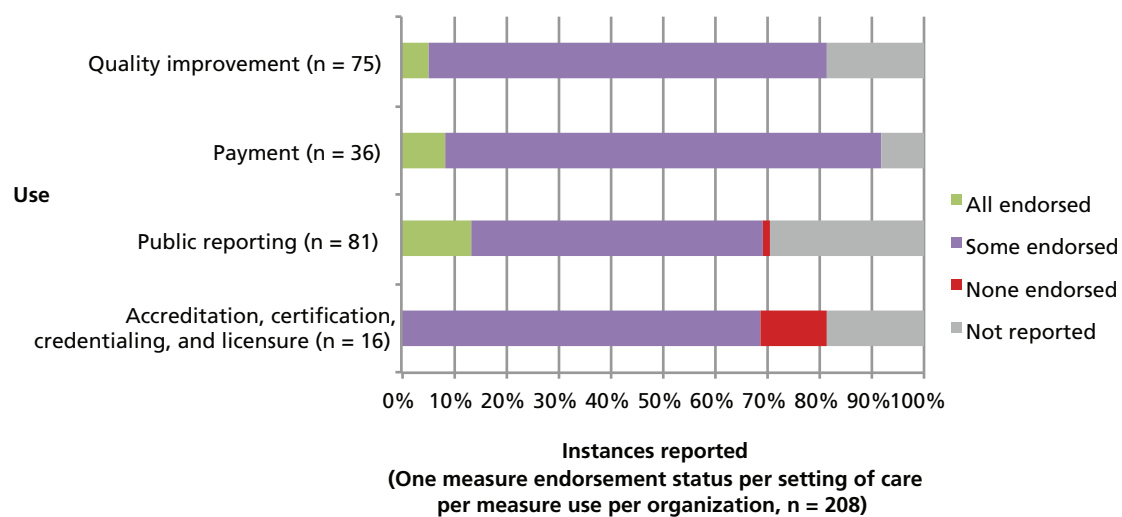**(One set of measure domains per setting of care per measure use per organization, n = 743)**

NOTE: The number to the right of each type of measure use shows the total number or instances of measure domains being reported. The bars show the percentage of instances involving each type of measure domain within each specific measure use.
RAND *TR1148-3.8*

tory setting focused heavily on Healthcare Effectiveness Data and Information Set (HEDIS) measures and the Consumer Assessment of Healthcare Providers and Systems (CAHPS) Clinician and Group Survey patient experience survey. For nursing home and home health settings, the most frequently used measures are those reported in the Centers for Medicare and Medicaid Services (CMS) Compare programs. Additional measures used by some organizations include nonendorsed cost and utilization measures and internally developed surveys and measures for QI. For example, one organization reported using a survey to generate physician ratings of other physicians. Another organization reported having more than 650 measures, some endorsed and some internally developed, which they use in the context of physician QI modules.

Providers That Are the Focus of Measurement Activities

Each organization could focus its measurement efforts on multiple types of providers and plans (i.e., the units of measurement), as shown in Figure 3.9 (see also Figure D.1 in Appendix D). Hospital performance was the most frequently measured, with 54 percent (68 of 127) of uses measuring them, and individual physicians and physician sites were measured somewhat less frequently (38 percent [48 of 127] and 28 percent [36 of 127] of measure uses, respectively).

Data Used to Construct Measures

Each organization could use multiple types of data sources for each measure use. In sum, the 70 organizations employed 261 data sources for 127 measure uses, or, on average, 2.1 data sources per use. Claims and administrative records were the most common source of data and were used in 76 percent (96 of 127) of identified measure uses. Patient surveys, the next most frequently used type of data, were included in 56 percent (71 of 127) of uses. To a greater extent than the other uses, accreditation, certification, credentialing, and licensure uses tended to include data from medical records and organization self-reports to construct measures (Figure 3.10 and Figure D.1 in Appendix D). Information from the key-informant interviews

Figure 3.9
Units Being Measured by Different Uses (n = 70 organizations)
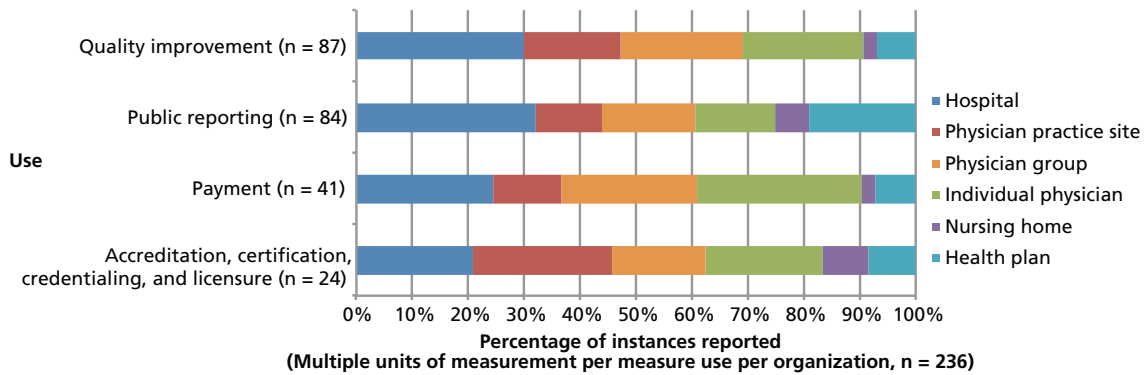
NOTE: The number to the right of each measure use shows the total count of measured provider types reported by end users for a given measure use. The bars show the percentage of instances involving each unit of measurement for each specific measure use.
RAND *TR1148-3.9*

found that claims and administrative data were the most frequently used source of information used to construct measures, but they also used medical records, surveys, and self-reported data.

Factors Driving Measure Use

According to our key-informant interviews, a combination of internal and external factors is driving the use of various measures. External factors include federal or state legislation mandating public reporting or value-based purchasing (VBP), such as the Patient Protection and Affordable Care Act of 2010. Some organizations are also responding to pressure from purchasers in their communities or nationally to improve quality and lower costs or to secure NCQA accreditation or recognition. Other organizations are utilizing measures as a part of their internal commitment to QI or to advance their stated mission. One interviewee, expressing the organization's purpose for using measures, stated, "We're trying to give consumers meaningful, actionable information so they can advocate for better care on their own behalf." Another organization commented, "The goal of our performance-measurement strategy is to advance care to our end-state vision of safe, affordable, effective, quality patient-centered care."

> *We're trying to give consumers meaningful, actionable information so they can advocate for better care on their own behalf.*

Question 2: What Factors Are Influencing Measure Use, Particularly Endorsed Measures?

Process Used to Select Measures

Most of our interviewees described an established set of criteria they utilize when selecting measures. There was a clear preference for endorsed measures for payment and public reporting purposes. However, there was also a willingness among most end users to use nonendorsed measures for priority areas in which endorsed measures were not available or feasible to use. One interviewee expressed, "NQF endorsement is very important. If there is not an endorsed measure and we can stand in front of providers and defend the measure, we would use it." The use of nonendorsed measures typically reflects a need either to modify measures to local

Figure 3.10
Data Sources for Different Measure Uses (n = 70 organizations)
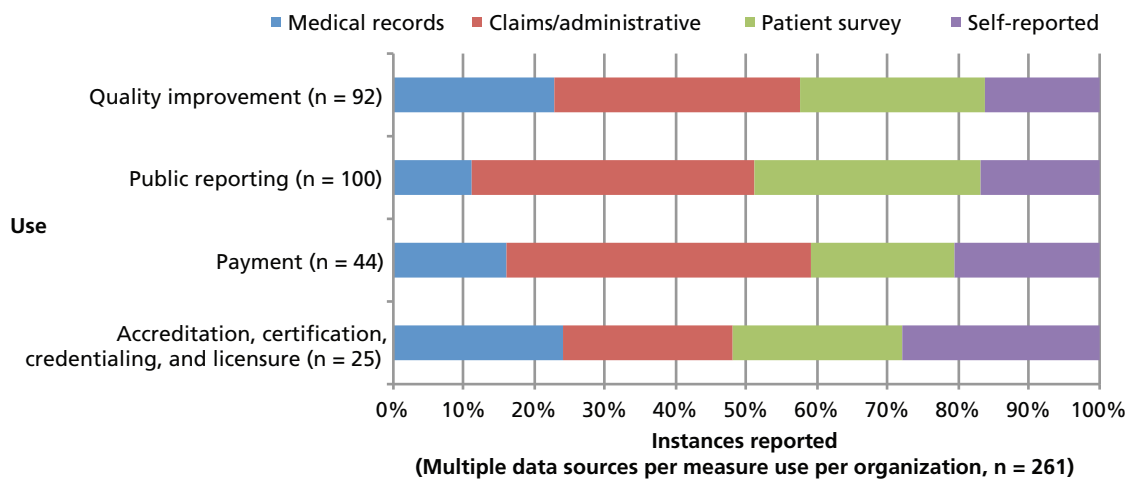
NOTE: The number to the right of each type of measure use shows a count of the total mentions of data sources (multiple data sources per measure use per organization). The bars shows what percent involved each type of data source for each type measure use.
RAND *TR1148-3.10*

> *NQF endorsement is very important. If there is not an endorsed measure and we can stand in front of providers and defend the measure, we would use it.*

circumstances (e.g., adapting the specification of a measure to utilize claim data instead of medical record data or vice versa) or to measure an area that is not yet represented in the NQF-endorsed set of measures (e.g., resource use or efficiency measures). Other criteria commonly applied in the selection process include relevance to members, consumers, or providers; feasibility of data collection; and scientific acceptability. For those engaged in QI, the ability to get data in real time is very important (i.e., data that represent the provider's current performance). For those organizations using measures for public reporting, the measures need to be understandable to their audiences. As a result, many organizations reported using composite measures, which they believe to be more easily interpreted than reporting numerous individual clinical measures.

Stakeholder input is also an important influence on measure selection and use. In fact, several organizations have a formal comment process in order to gain input on proposed measures. One organization has each of its specialist groups develop performance measures for which it will be accountable. Stakeholder input regarding the burden of data collection also had an impact on the number and types of measures used.

Measure selection was described as a continuous activity in which organizations adjust the measure set that they are using based on changes in the evidence base, changes in performance (i.e., "topping out" or perceived weakness in a measure), data availability, provider pushback, evidence of unintended consequences, and other factors.

Factors Facilitating or Impeding Use of Measures

The availability of data to construct measures, obtained either from providers or from public databases, was the single most important factor cited as either facilitating or impeding the

use of measures. Having a strong data infrastructure to allow the construction of measures (e.g., multipayer, timely data, clinically enriched data) allows more opportunities for measure construction and use. Some organizations noted the prospect of more clinically enriched data becoming available moving forward with the advent of electronic health records (EHRs) but added that EHRs are not sufficiently robust at this stage to generate this information for measure construction. An interviewee from one organization that collects data directly from providers observed, "Ambulatory outcome measures that draw from clinical data require populating an Excel spreadsheet at this point in time because EHRs are not yet built [to] do this and [EHRs] do not talk to each other." Conversely, another organization noted, "Data infrastructure is key—physician groups with good [information technology] systems are able to use measures much better." The problem of varying data infrastructure across facilities was mentioned by one interviewee, which made standardization of measure implementation and use difficult. Still another impediment is the lack of availability of population-level data, such as the number of people in a region who have diabetes.

> *The data infrastructure is key. Physician groups with good information technology systems are able to use measures much better.*

Provider trust in both the process used to measure provider performance and in the underlying evidence base of the measures for which providers are being held accountable is another factor that serves to facilitate measure use. In most cases, it was noted that NQF endorsement or widespread use of a measure enhanced provider buy-in. It was also observed that physicians tend to trust the national registries because they understand the data being collected and have confidence in the specialty organizations that operate these registries. One interviewee stated that, in her experience, "Resistance to measures can be quickly dealt with because the measures used are validated and peer-reviewed." Another organization experienced a great deal of provider resistance to using a CAHPS survey in its P4P program prior to the receipt of NQF endorsement. However, several individuals observed that many physicians do not understand the rigor of the NQF process and therefore are not swayed by endorsement.

The alignment of measures across reporting initiatives is a key facilitator of measure use because it substantially reduces the reporting burden on providers. Conversely, lack of alignment is a major impediment. One interviewee observed that hospital administrators are challenged by the array of demands from different end users. The terms "measure fatigue" and "registry fatigue" were both used to describe the current situation. Several interviewees also raised the issue of cost in responding to the various measure collection requirements: "Issues that organizations flag include having to meet the demands of multiple requirements for measures [e.g., HHS, the states, purchasers, and payers], and this is a cost burden to collect measures in multiple different ways to satisfy all these different end users." Some organizations have made an effort to align the measures they are using with national, state, or regional programs to address these issues. One individual observed, "the National Quality Strategy might encourage alignment among various users going forward."[1]

> *It is a cost burden to collect measures in multiple different ways to satisfy all these different end users.*

In some cases, provider training on how to abstract data and education about evidence supporting measures

---

[1]   The National Quality Strategy is a federal effort to create national aims and priorities to guide local, state, and national efforts to improve the quality of health care in the United States.

has facilitated their use. One organization has worked with providers on using measures for QI as a way to boost their performance, as well as to prepare them for potential public reporting and P4P initiatives.

The National Quality Strategy might encourage alignment among various users going forward.

Another major impediment to the use of measures is a lack of timely data to construct measures (i.e., preferably data within the past year of performance). This is a particular problem for those organizations that rely on public data sources. Examples of data that are not timely, according to interviewees, include central-line infection data, which typically come from the states, and HCAHPS data, which come from Hospital Compare.

Those organizations that are measuring performance at the individual physician or physician group level often cited the problem of "small numbers," particularly for HEDIS measures.[2] As one person noted, "Due to small numbers, what works at the health plan level doesn't work at the doctor's office level." One individual explained that his organization addressed the problem of small numbers by using a two-year rolling average for HEDIS measures. Another respondent expressed, "For many of the measures, the population is so small that [the measures] aren't usable, even at the group level. It may not make sense to endorse these."

Due to small numbers, what works at the health plan level doesn't work at the doctor's office level.

An additional impediment is the selection of measures, given the sheer number available. One organization described the list of NQF-endorsed measures as "endless" and making prioritization very difficult, particularly among ones that are similar, such as the appropriate level of HbA1C for diabetic patients.

For many of the measures, the population is so small that they aren't usable, even at the group level. It may not make sense to endorse these.

Anticipated Changes to Measures Used Moving Forward

Many respondents expressed the hope that increased adoption of registries and EHRs (with expanded capabilities) would provide greater capacity for broader and different types of measures, such as clinical outcomes. Some individuals were not convinced that EHRs would have much impact in the near- to mid-term time horizon due to a lack of interoperability, variability in the functionalities, and the fact that "they don't aggregate data in a meaningful way." One interviewee commented, "measures must be built into electronic health record systems and the work flows instead of [included as] an afterthought."

Interviewees also expressed that they would continue to adjust and evolve the set of measures they are using for different applications. There was interest expressed in incorporating new types of measures, such as longitudinal outcomes, patient functional status, cost and resource use, and additional patient safety measures. There was also interest in being able to use measures to assess performance in new settings, such as nursing homes and home health, and for new populations, such as patients who are dually eligible for Medicare and Medicaid. Last, new applications, such as VBP, accountable care organizations (ACOs), bundled payments, and other payment and system reforms will necessitate new or different applications of existing

---

[2] Here, the term *small-numbers problem* refers to too small of a sample of patients to generate a reliable estimate of performance (i.e., sufficient signal rather than noise in the estimate).

measures or new measures. This might mean, for example, measures that assess coordination of care when patients move between settings of care in bundled payment applications.

Question 3: What Types of System Changes Have Occurred as a Result of Measure Use?

Although many improvements were noted, few organizations had done formal evaluations to document their results, and even fewer had attempted any type of return-on-investment (ROI) analysis. Four organizations reported published peer-reviewed articles, and two organizations included an ROI analysis in those studies (Chernew et al., 2011; Song et al., 2011; Drenkard, 2010; Chassin et al., 2010; Hibbard, Stockard, and Tusler, 2005). Respondents mentioned challenges associated with conducting formal studies, such as a lack of control groups, difficulty in measuring costs for ROI analysis, and the resources required to do evaluation. Several said that they are just beginning to assess the impacts in a more formal way.

However, interviewees anecdotally reported a range of system changes in cost, quality, patient safety, and patient experience resulting from their use of measures. They noted, although they did not provide actual numbers, that they had realized cost savings from initiatives to increase the use of generic prescribing and from implementation of tiered health plan products, in which the tiering was based on use of performance measures. QIs, such as a dramatic increase in compliance on high-stakes measures and changes to physician practices as a result of an MOC QI program, were among the improvements mentioned. Additionally, an interviewee from one organization expressed the belief that its public reporting has "been helpful in getting hospitals to prioritize things like CPOE [computerized physician order entry]." An interviewee from another organization described an improvement in CAHPS scores following the inclusion of the survey results in its P4P program.

On the other hand, a small number of interviewees questioned the impact of their measurement efforts. Those who are reporting quality information to consumers are not able to track how consumers use the information and are skeptical about how much weight the information currently carries in consumer decisionmaking. Several interviewees questioned how much effect measures have had on outcomes, especially in addressing chronic conditions. One organization noted that physician performance peaked in 2008 and has since declined somewhat.

Question 4: What Improvement Areas Could Facilitate the Use of Standardized Endorsed Performance Measures by Various End Users?

Measure Gaps and Opportunities

Measurement gaps or opportunities can be characterized by the type of measure, the condition to be measured, or the setting or population to be measured. One of the most frequently mentioned measure gaps was coordination of care. One interviewee characterized 30-day readmission to a hospital as a "gross measure" of care coordination and suggested two alternative process measures: (1) patients who are discharged with "clear, understandable care plans" and (2) patients who have a physician visit within five to seven days after discharge. One respondent felt that generating performance scores on these types of measures would "depend

on better databases" and commented that, instead, the respondent's organization is using patient survey questions, such as "Were your tests available when you went to the doctor?" Interviewees would also like to see more and different types of outcome measures developed and endorsed. This includes additional measures of complications and patient safety (inpatient and ambulatory) but also some newer types, such as longitudinal measures, those that consider a patient's risk profile, and measures of patient-reported health or functional status that could be used across conditions.

Additional types of measure gaps identified were related to cost and resource use, such as the per capita cost per type of service and commercial bed days per thousand; nursing-sensitive processes that affect outcomes, such as assessment, evaluation, and planning; population health indicators for ACOs and communities; methodologically sound composites; measures of shared decisionmaking; and access and timeliness indicators. One interviewee commented on the issue of access: "The only measure that comes close is a CAHPS question, but it is only asked of people with insurance who are continuously enrolled for one year, which is a problem for Medicaid [recipients]."

One condition for which many of those interviewed felt that there is a paucity of measures is pregnancy. Other condition-specific measure gaps that interviewees mentioned included obesity, end-of-life care, mental illness, spinal-cord and traumatic brain injuries, heart failure (inpatient), and some specialties (dermatology, ophthalmology, gastroenterology, infectious disease, and critical care). However, some respondents felt that limited resources should not be used to develop measures for every specialty. They said that measure development should focus on identifying areas that are applicable across a range of conditions touching many different specialties. Some of the interviewees expressed some concern that existing specialty-specific measures are too easy and do not require much change on the part of the provider to succeed on the measures.

Several different settings and populations for which gaps exist were also identified. The settings mentioned were nursing homes, home health, assisted living, and schools for nurse management of chronic conditions. The populations mentioned were dual eligibles and adolescents.

Advice to Improve the Uptake of Standardized Measures
The expression of need for more measures was counterbalanced by comments regarding the need for better alignment and prioritization of existing measures to help in selection and use of measures. One interviewee commented, "It is less a demand for more measures and more a problem that the extant measures are very distracting and not aligned." Interviewees expressed a need to ease end-user selection of measures by "weeding out" or rationalizing endorsed measures or selecting the "best in class" when there are very similar endorsed measures. Similarly, one organization expressed a desire for guidance on choosing the most important and effective measures to drive change. Other comments reflected similar concerns, including these:

It is less a demand for more measures and more a problem that the extant measures are very distracting and not aligned.

> We should simplify and re-evaluate the goals of measurement. We need to be clear about what we want out of our system and measure that.

The measure endorsement process is driven more by what is easily measurable rather than by what is important and is focused on physician rating/selection, which is generally impractical or worse, rather than improving quality.

Endorsed measures should be a small set of high-leverage, patient-oriented outcome measures.

There were specific opportunities directed at NQF to create end-user support tools and technical assistance. In particular, end users suggested improvements to the NQF website and search engine to allow end users to search measures by the population of interest or the data source, more easily obtain measure specifications, and obtain information about where and how a particular measure is being used. Another suggestion was for NQF to advance the conversation about composite use given the different approaches currently being deployed. One individual suggested that NQF, possibly in partnership with the Agency for Healthcare Research and Quality, provide a clearinghouse for QI measures. He observed,

There is a secondary market for measures that organizations have developed internally for QI, but [there is] no incentive for them to submit them for endorsement. The NQF measures tend to be all for accountability and public reporting but are largely feckless for improvement work.

Outreach and education represents an area of opportunity for end-user support by NQF. In an effort to gain a common understanding of performance measurement and the endorsement process, some suggested that NQF offer stakeholder training, potentially with continuing medical education credit. One organization commented, "NQF should publicize the rigor behind measures. One place to start is with physicians since they often do not know about or understand these initiatives." There was a suggestion that NQF do "road shows" to "facilitate the work going on in communities" and participate in community collaboratives that are testing and reporting measures. Interviewees also suggested that physicians and hospitals could be engaged through their various specialty societies and trade groups.

NQF should publicize the rigor behind measures.

There was some disagreement about NQF's potential role vis-à-vis measure development. One interviewee encouraged more involvement in development, saying, "I recognize [that] NQF is in a tough position as an endorser and convener, but not a developer. [It] should get more hands-on in the process of development and testing." Another interviewee felt that NQF should be in "testing, validating, and demonstrating clear specifications mode" rather than generating more measures.

Although many of the individuals with whom we spoke praised NQF and its efforts, some felt that the endorsement process could be streamlined and improved. The process was viewed by some as "onerous" and an "impediment" to progress and a way for providers to "slow down measure development." This was countered by the comment that "Keeping the approval process rigorous is important. Don't try to fill gaps by making it easier to put measures through." One individual hoped that NQF would better publicize upcoming calls for measures and the review calendar so that organizations could coordinate their measure-review processes with NQF's. Another interviewee suggested that it would be helpful for NQF to pro-

vide more advanced notice of upcoming calls for measures so that measure developers would have more time to test them prior to submission.

Conclusions

Future Opportunities to Support the Use of Measures

As the measurement development, endorsement, and application communities reflect on the results of this study, we note that several issues emerged that represent opportunities and potential areas of focus moving forward to help facilitate the use of measures:

- *Establish priorities for where end users should focus their attention and resources.* Although NQF has a large number of endorsed measures, the sheer number is daunting to end users, and many measures are viewed as irrelevant or low impact. Given the limited resources and attention that end users can devote to measurement, they need guidance on defining a more limited set of high-value, high-leverage measures to which they should devote resources that will lead to real improvements in outcomes or efficiency (fewer resources spent without reductions in quality). Outcome measures were seen as high-value measures that should be prioritized.
- *Align measures.* Among those who are the focus of measurement (i.e., providers), there is an important need to help reduce the burden of reporting to a wide array of entities by ensuring that the measures used by the entities are aligned. Currently, reporting efforts among the various stakeholders operate independently, are frequently not aligned, and lead to increased data collection, analysis, and reporting burdens for providers. Establishing measure priorities can create an opportunity for better alignment among the various parties that impose reporting requirements. Cataloging the specific types of measures being used for each end use could help inform efforts to align measures.
- *Develop new measures for new measure uses.* The measure-use landscape continues to evolve, particularly in response to health system reform efforts. End users are preparing for new measure-use opportunities and the need for different types of measures, particularly in the payment use area (e.g., ACOs, VBP applications, episode-of-care and bundled payment models). Additionally, there is a desire for crosscutting measures that apply to all providers both to broaden the group of providers that can be assessed and to minimize the number of measures that have to be constructed. Additionally, end users expressed strong interest in new measures that would track the functioning of the patient and how functioning and other key outcome markers change longitudinally (i.e., improvement measures).
- *Build support for the use of measures.* The criteria that are most salient to end users in choosing measures for use are (1) the measures are relevant to providers who must act on them, as well as consumers and payers; (2) the data needed to construct the measure are

feasible to collect; and (3) the measure has a scientific evidence base to ensure its accept-ability to providers.

- *Measure construction requirements in the evolving data landscape.* At present, end users are relying heavily on claims and administrative data and patient surveys and are not yet able to systematically and efficiently capture information that will enable the construc-tion of intermediate and longitudinal outcomes. At a minimum, a limited, prioritized set of measures should be the focus of front-end planning to influence the data architecture (e.g., data fields and not free text) of EHRs and other HIT to support the construction of measures. The development of better data sources, such as what is envisioned through HIT sources, has the potential to enable better measurement and use of measures.

- *Conduct a systematic review of the literature to fully catalog measure use.* The scope of this project did not permit a full review of the literature to examine uses of measures for a variety of purposes and to explore the secondary use of measures (e.g., by providers who are exposed to performance results in public report cards, consumers who are shown performance report cards). A thorough review of the literature would extend the work in this study and provide for a deeper understanding of uses and issues related to use of measures.

- *Formally assess the system-change results from the use of measures.* Although end users cite anecdotal examples of the benefits that have accrued from use of performance measures, it is unclear how many of these benefits have been formally evaluated and documented. This underscores the need for a future study that would systematically attempt to quan-tify the impacts that have resulted from measure use, given the substantial investment that has been made to use measures to drive system change.

- *Create support tools to help end users.* End users want access to full measure specifications, more information about where and how measures are being used, and an ability to more easily search the measures contained in the NQF-endorsed set of measures. Improved search functions and new tools could facilitate end users' ability to use measures. End users also see value in increased outreach and education to stakeholders in their com-munities (particularly physicians and other health-care providers) about the validity of the measures and why measurement is important as a means of better engaging them in performance improvement.

Members of the Evaluation of Measure Use Expert Advisory Panel

Table A.1
Members of the Evaluation of Measure Use Expert Advisory Panel

Name	Organization
Lisa Backus	Veterans Health Administration
Clarence Braddock	Stanford University
David Bundy	Johns Hopkins University School of Medicine
Carol Cronin	Informed Patient Institute
Michael Doering	Pennsylvania Patient Safety Authority
Joseph Francis	Veterans Health Administration
Kate Goodrich	Office of the Assistant Secretary for Planning and Evaluation, HHS
David Hoyt	American College of Surgeons
Thomas James	Humana
Alyssa Keefe	Federal Regulatory Affairs, California Hospital Association
Keith Krein	Kindred Healthcare
Diane Mayberry	Minnesota Community Measurement
Richard Morin	Mayo Clinic
Rita Munley Gallagher	American Nurses Association
David Nau	Pharmacy Quality Alliance
Devorah Rich	Greater Detroit Area Health Council
Joachim Roski	Brookings Institution
Jonathan Sugarman	Qualis Health
Phyllis Torda	NCQA

List of Organizations Included in the Document and Website Review

Following are the organizations included in our document and website review:

Aetna	The Alliance
American Academy of Family Practice	American Academy of Pediatrics
American Board of Internal Medicine	American College of Cardiology
American College of Physicians	American Nurses Credentialing Center
Anthem Blue Cross	Blue Cross Blue Shield of Massachusetts
Blue Shield of California	Buyers Health Care Action Group
California Cooperative Healthcare Reporting Initiative	California Office of the Public Advocate, Department of Managed Health Care
California Public Employees' Retirement System (CalPERS)	California Quality Collaborative
Care Transitions Quality Improvement Organization Support Center	Centers for Disease Control and Prevention
CMS	Colorado (Medicaid)
Commonwealth Fund (*Why Not the Best*)	Community Health Accreditation Program
Consumer Checkbook	Consumers Union (*Consumer Reports*)
Employers' Coalition on Health	Florida Hospital Association
Group Health of Puget Sound	Harvard Pilgrim
Health Partners	HealthPlus
Health Resources and Services Administration for Federally Qualified Health Centers	Healthy Memphis Common Table
Highmark	Hill Physicians Medical Group
Horizon/Blue Cross Blue Shield of New Jersey	Humbolt del Norte Medical Group

Institute for Healthcare Improvement	Integrated Healthcare Association
Intermountain Healthcare	Iowa Healthcare Collaborative
Joint Commission	Kaiser Permanente
L.A. Care Health Plan (Medicaid)	Leapfrog Group
Louisiana Department of Health and Hospitals	Maine Health Management Coalition
Maryland Health Care Commission	Massachusetts Group Insurance Commission
Mayo Clinic	Minnesota Medicaid
Massachusetts Health Quality Partners	Minnesota Community Measurement
National Business Coalition on Health (eValue 8)	NCQA
Nevada Health Care Coalition	New York Medicaid
New York Quality Alliance	Oklahoma Foundation for Medical Quality
Premier Hospitals	Puget Sound Health Alliance
Quality Quest for Health of Illinois	Society for Thoracic Surgeons
St. Louis Area Business Health Coalition	United Healthcare
URAC	*U.S. News and World Report*
Vermont Department of Banking, Insurance, Securities and Health Care Administration	Veterans Health Administration
Virginia (Medicaid)	Washington (Medicaid)

Breakdown of Measure Uses in Sample of Organizations Undergoing Document and Website Review

Figure C.1 summarizes the structure of the data collected through the website and document reviews, the results of which are described in this report. Each organization was categorized as one of 11 organization types and could report measures for up to four measure uses (QI, public reporting, payment, and accreditation, certification, credentialing, and licensure). We identified a total of 127 measures uses across the 70 organizations. For each measure use, we recorded its geographic scope (e.g., state, national), data sources (e.g., claim data, medical records, surveys), settings of care (e.g., inpatient, ambulatory, nursing home), and units of measurement (e.g., hospital, physician group, individual physician). Although each measure use had a single geographic scope, it could use multiple data sources, focus on multiple units of measurement, and include multiple care settings. For each measure use in each setting, we also recoded the set of measure domains included (e.g., process, outcome, patient experience) and its NQF endorsement statuses. Each setting's measure domains had a single NQF endorsement status (all endorsed, some endorsed, none endorsed, unable to determine).

Figure C.1
Structure of Website and Document Review Data
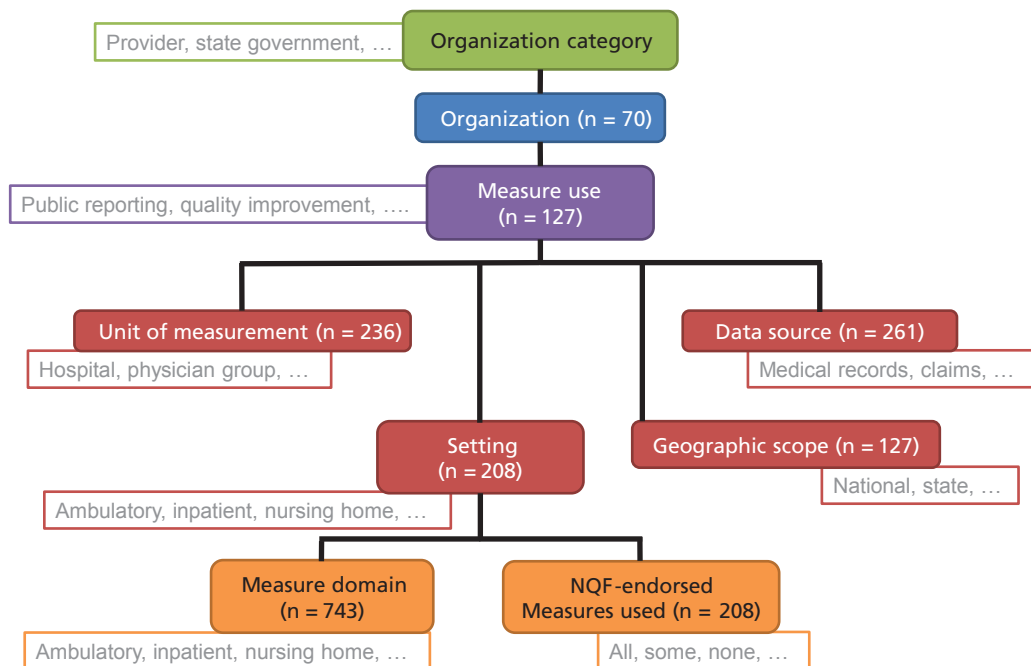
RAND *TR1148-C.1*

To illustrate the data structure, consider an example, Hypothetical Health Organization (HHO). Suppose HHO is a provider organization, and it conducts two measure uses: public reporting and QI. Each of these uses has a geographic scope: a statewide public reporting site and a local network QI initiative. Each of HHO's uses also has associated data sources; perhaps the public reporting site uses claims and patient survey data and the QI initiative uses medical records. The uses also have units of measurement; for example, the public reporting site could provide information about hospitals and physician practice sites. Each use has a set of settings of care: The public reporting has inpatient and ER measures and the QI has ambulatory measures. Each of those settings has a set of measure domains, and the set of domains has an overall NQF endorsement status. For example, say that the HHO QI initiative outpatient setting uses process, structure, and patient-experience measures, only some of which are NQF endorsed.

Additional Data Summary Tables from Document and Website Review and Key-Informant Interviews

Table D.1
Characteristic of Measurement Activities by Measure Use (n = 70 organizations)

Characteristic	Accreditation, Certification, Credentialing, and Licensure (n = 11)	Payment (n = 23)	Public Reporting (n = 45)	QI (n = 48)
Data source				
Medical records	6 (54.5%)	7 (30.4%)	11 (24.4%)	21 (43.8%)
Claims/ administrative	6 (54.5%)	19 (82.6%)	40 (88.9%)	32 (66.7%)
Patient survey	6 (54.5%)	9 (39.1%)	32 (71.1%)	24 (50%)
Self-reported	7 (63.6%)	9 (39.1%)	17 (37.8%)	15 (31.2%)
Unit of measure				
Hospital	5 (45.5%)	10 (43.5%)	27 (60%)	26 (54.2%)
Physician practice site	6 (54.5%)	5 (21.7%)	10 (22.2%)	15 (31.2%)
Physician group	4 (36.4%)	10 (43.5%)	14 (31.1%)	19 (39.6%)
Individual physician	5 (45.5%)	12 (52.2%)	12 (26.7%)	19 (39.6%)
Nursing home	2 (18.2%)	1 (4.3%)	5 (11.1%)	2 (4.2%)
Health plan	2 (18.2%)	3 (13%)	16 (35.6%)	6 (12.5%)

NOTE: n = the number of organizations reporting that end use. Within each column set, the percentage is the number of organizations that reported that activity divided by N organizations using measures for that purpose. For example, of the 11 organizations using measures for accreditation, certification, credentialing, or licensure, six, or 54.5 percent of the 11, are using medical records.

Table D.2
Characteristic of Measurement Activities by Measure Use (n = 70 organizations)

Characteristic	Accreditation, Certification, Credentialing, and Licensure (n = 16)	Payment (n = 36)	Public Reporting (n = 81)	QI (n = 75)
Setting of care				
Inpatient	6 (37.5%)	11 (30.6%)	27 (33.3%)	26 (34.7%)
Ambulatory	8 (50%)	17 (47.2%)	37 (45.7%)	31 (41.3%)
ER	1 (6.2%)	5 (13.9%)	8 (9.9%)	12 (16%)
Nursing home	0 (0%)	1 (2.8%)	6 (7.4%)	3 (4%)
Home care	1 (6.2%)	2 (5.6%)	3 (3.7%)	3 (4%)
Measure domain				
Structure	5 (31.2%)	8 (22.2%)	16 (19.8%)	14 (18.7%)
Process	8 (50%)	19 (52.8%)	39 (48.1%)	40 (53.3%)
Outcome	7 (43.8%)	13 (36.1%)	38 (46.9%)	35 (46.7%)
Access	2 (12.5%)	3 (8.3%)	15 (18.5%)	10 (13.3%)
Safety	2 (12.5%)	6 (16.7%)	25 (30.9%)	21 (28%)
Cost	3 (18.8%)	12 (33.3%)	22 (27.2%)	18 (24%)
Patient experience	5 (31.2%)	11 (30.6%)	33 (40.7%)	23 (30.7%)
NQF endorsement				
All endorsed	0 (0%)	3 (8.3%)	11 (13.6%)	4 (5.3%)
Some endorsed	11 (68.8%)	30 (83.3%)	45 (55.6%)	57 (76%)
None endorsed	2 (12.5%)	0 (0%)	1 (1.2%)	0 (0%)
Not reported	3 (18.8%)	3 (8.3%)	24 (29.6%)	14 (18.7%)

NOTE: Organizations could be using measures for the same end use in multiple settings (e.g., accreditation for inpatient and nursing-home settings, which would count as two instances of use within accreditation, certification, credentialing, and licensure). So, of the 11 organizations reporting use of measures for accreditation, certification, credentialing, or licensure, there were 16 instances of use covering multiple settings. The counts (n) in this table account for use within setting.

Table D.3
Characteristics of Measurement Activities by Organization Type

Characteristic	Accreditation, Certification, Credentialing, and Licensure Bodies (n = 5)	Community Collaboratives (n = 10)	Consumer Groups (n = 2)	Federal Government (n = 4)	Health-Care Improvement Organizations (n = 4)	Other (n = 2)	Payers (plans) (n = 11)	Provider Specialty Organizations (n = 5)	Providers (n = 7)	Purchasers and Purchaser Coalitions (n = 10)	State Government (n = 8)
Geographic scope[a]											
National	5	0	2	9	1	3	0	10	1	1	0
State	0	10	0	0	4	0	0	0	2	4	12
Network	1	1	0	0	0	0	17	0	12	3	0
Other	4	2	0	0	2	0	2	0	0	2	0
Data source[a]											
Medical records	3	1	1	7	7	0	8	5	10	2	1
Claims/administrative	5	12	2	6	5	3	19	6	15	11	11
Patient survey	7	8	2	6	2	2	11	4	12	7	7
Self-reported	10	6	0	4	1	1	7	7	0	6	3
Setting of care[a]											
Inpatient	6	7	2	6	6	3	12	5	10	8	7
Ambulatory	8	11	1	7	4	1	18	5	12	14	10
ER	0	6	0	5	1	0	8	0	5	2	3
Nursing home	1	0	0	2	0	2	5	0	0	0	2
Home care	1	0	0	1	0	1	3	0	3	0	1

Table D.3—Continued

Characteristic	Accreditation, Certification, Credentialing, and Licensure Bodies (n = 5)	Community Collaboratives (n = 10)	Consumer Groups (n = 2)	Federal Government (n = 4)	Health-Care Improvement Organizations (n = 4)	Other (n = 2)	Payers (plans) (n = 11)	Provider Specialty Organizations (n = 5)	Providers (n = 7)	Purchasers and Purchaser Coalitions (n = 10)	State Government (n = 8)
Unit of measure[a]											
Hospital	4	5	2	6	6	3	14	7	10	8	5
Physician practice site	7	2	0	6	0	0	7	3	5	2	2
Physician group	3	9	1	2	2	0	11	5	9	2	0
Individual physician	2	4	1	5	0	0	19	3	5	5	2
Nursing home	3	0	0	2	0	1	3	0	0	0	2
Health plan	4	3	0	1	1	1	4	0	0	4	9
Domain of measure[b]											
Structure	2	3	0	5	1	1	12	7	3	6	3
Process	8	12	1	8	7	2	19	10	13	12	10
Outcome	7	10	2	7	3	3	13	8	14	12	7
Access	4	3	0	2	1	1	8	0	2	4	3
Safety	5	3	2	7	5	3	12	0	9	6	3
Cost	5	8	0	5	1	2	8	0	11	8	4
Patient experience	7	9	2	5	1	2	12	4	12	9	7

Characteristic	Accreditation, Certification, Credentialing, and Licensure Bodies (n = 5)	Community Collaboratives (n = 10)	Consumer Groups (n = 2)	Federal Government (n = 4)	Health-Care Improvement Organizations (n = 4)	Other (n = 2)	Payers (plans) (n = 11)	Provider Specialty Organizations (n = 5)	Providers (n = 7)	Purchasers and Purchaser Coalitions (n = 10)	State Government (n = 8)
NQF endorsement status[b]											
All endorsed	4	10	0	4	4	2	3	0	1	3	9
Some endorsed	9	10	1	11	6	4	36	6	25	11	4
None endorsed	2	0	0	0	0	0	0	0	0	1	0
Not reported	1	4	2	6	1	1	7	4	4	9	10

[a] The units are at the level of the measure use within the organization.

[b] The units are at the level of the setting of care of the measure use within the organization.

Table D.4
Measure Domains Used, by Setting of Care That Was the Focus of Measurement
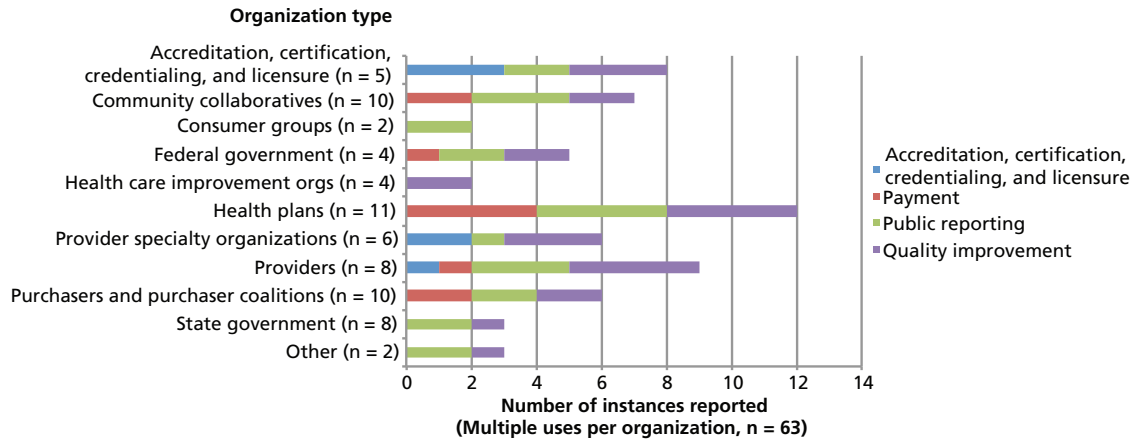**(n = 70 organizations)**

Measure Domain	Inpatient (70 instances)	Outpatient (93 instances)	ER (26 instances)	Nursing Home (10 instances)	Home Care (9 instances)
Structure	25 (35.7%)	31 (34.1%)	7 (36.9%)	4 (40%)	2 (22.2%)
Process	65 (92.9%)	82 (90.1%)	26 (100%)	6 (60%)	7 (77.8%)
Outcome	51 (72.9%)	62 (68.1%)	11 (42.3%)	6 (60%)	6 (66.7%)
Access	3 (4.3%)	28 (30.8%)	3 (11.5%)	2 (20%)	2 (22.2%)
Patient experience	45 (64.3%)	51 (56%)	14 (53.8%)	3 (30%)	5 (55.6%)
Safety	48 (68.6%)	22 (24.2%)	15 (57.7%)	7 (70%)	8 (88.9%)
Cost	26 (37.1%)	48 (52.7%)	11 (42.3%)	4 (40%)	7 (77.8%)

Table D.5
Number of Uses, by Organization Type (n = 70 Organizations)

Organization Type	Accreditation, Certification, Credentialing, and Licensure	Payment	Public Reporting	QI	Total Uses
Accreditation, certification, credentialing, and licensure bodies	5	0	2	3	10
Community collaboratives	0	3	9	5	17
Consumer groups	0	0	2	0	2
Federal government	1	1	4	4	10
Health-care improvement organizations	0	0	0	4	4
Payers (plans)	0	9	7	10	26
Provider specialty organizations	4	0	2	6	12
Providers	1	3	5	8	17
Purchasers and purchaser coalitions	0	5	6	3	14
State government	0	2	6	4	12
Other	0	0	2	1	3

[a] Represents the number of organizations that reported using measures for that purpose.

Figure D.1
Measure Use, by Organization Category, Interview Sample (n = 30)

Organization type

Legend:
- Accreditation, certification, credentialing, and licensure
- Payment
- Public reporting
- Quality improvement

X-axis: **Number of instances reported**
(Multiple uses per organization, n = 63)

Y-axis categories:
- Accreditation, certification, credentialing, and licensure (n = 5)
- Community collaboratives (n = 10)
- Consumer groups (n = 2)
- Federal government (n = 4)
- Health care improvement orgs (n = 4)
- Health plans (n = 11)
- Provider specialty organizations (n = 6)
- Providers (n = 8)
- Purchasers and purchaser coalitions (n = 10)
- State government (n = 8)
- Other (n = 2)

References

Agency for Healthcare Research and Quality, "Patient Safety Indicators Overview," undated. As of October 24, 2011:
http://www.qualityindicators.ahrq.gov/modules/psi_overview.aspx

Booz Allen Hamilton, *Synthesis of Evidence Related to 20 High Priority Conditions and Environmental Scan of Performance Measures: Final Report*, Washington, D.C.: National Quality Forum, January 13, 2010.

Chassin, Mark R., Jerod M. Loeb, Stephen P. Schmaltz, and Robert M. Wachter, "Accountability Measures: Using Measurement to Promote Quality Improvement," *New England Journal of Medicine*, Vol. 363, No. 7, August 12, 2010, pp. 683–688.

Chernew, Michael E., Robert E. Mechanic, Bruce E. Landon, and Dana Gelb Safran, "Private-Payer Innovation in Massachusetts: The 'Alternative Quality Contract,'" *Health Affairs*, Vol. 30, No. 1, January 2011, pp. 51–61.

Drenkard, K., "The Business Case for Magnet," *Journal of Nursing Administration*, Vol. 40, No. 6, June 2010, pp. 263–271.

HCAHPS—*See* Hospital Consumer Assessment of Healthcare Providers and Systems.

HHS—*See* U.S. Department of Health and Human Services.

Hibbard, Judith H., Jean Stockard, and Martin Tusler, "Hospital Performance Reports: Impact on Quality, Market Share, and Reputation," *Health Affairs*, Vol. 24, No. 4, July 2005, pp. 1150–1160.

Highsmith, Nikki, Roopa Mahadevan, and Alice Lind, *Building Bridges Among the National Quality Forum, Medicaid, and Other State Stakeholders*, Washington, D.C.: National Quality Forum, February 7, 2011.

Hospital Consumer Assessment of Healthcare Providers and Systems, undated home page. As of October 24, 2011:
http://www.hcahpsonline.org/home.aspx

Institute of Medicine, Committee on Quality of Health Care in America, *Crossing the Quality Chasm: A New Health System for the 21st Century*, Washington, D.C.: National Academy Press, 2001. As of October 24, 2011:
http://www.nap.edu/catalog/10027.html

———, Committee on Redesigning Health Insurance Performance Measures, Payment, and Performance Improvement Programs, *Performance Measurement: Accelerating Improvement*, Washington, D.C.: National Academies Press, 2006. As of October 24, 2011:
http://www.nap.edu/catalog/11517.html

Madison, Kristin, "The Law and Policy of Health Care Quality Reporting," *Campbell Law Review*, Vol. 31, No. 2, 2009, pp. 215–256.

McGlynn, Elizabeth A., "Introduction and Overview of the Conceptual Framework for a National Quality Measurement and Reporting System," *Medical Care*, Vol. 41, No. 1, January 2003, pp. I1–I7.

National Committee for Quality Assurance, *The State of Health Care Quality: Reform, the Quality Agenda and Resource Use*, Washington, D.C., 2010. As of October 24, 2011:
http://www.ncqa.org/portals/0/state%20of%20health%20care/2010/sohc%202010%20-%20full2.pdf

O'Neil, So, John Schurrer, and Sam Simon, *Environmental Scan of Public Reporting Programs and Analysis: Final Report*, Washington, D.C.: National Quality Forum, September 30, 2010.

Public Law 110-275, Medicare Improvements for Patients and Providers Act of 2008, July 15, 2008. As of October 24, 2011:
http://www.gpo.gov/fdsys/pkg/PLAW-110publ275/content-detail.html

Public Law 111-148, Patient Protection and Affordable Care Act, March 23, 2010. As of October 24, 2011:
http://www.gpo.gov/fdsys/pkg/PLAW-111publ148/content-detail.html

Safran, Dana Gelb, Melinda Karp, Kathryn Coltin, Hong Chang, Angela Li, John Ogren, and William H. Rogers, "Measuring Patients' Experiences with Individual Primary Care Physicians: Results of a Statewide Demonstration Project," *Journal of General Internal Medicine*, Vol. 21, No. 1, January 2006, pp. 13–21.

Song, Zirui, Dana Gelb Safran, Bruce E. Landon, Yulei He, Randall P. Ellis, Robert E. Mechanic, Matthew P. Day, and Michael E. Chernew, "Health Care Spending and Quality in Year 1 of the Alternative Quality Contract," *New England Journal of Medicine*, Vol. 365, No. 10, September 8, 2011, pp. 909–918.

U.S. Department of Health and Human Services, "Hospital Compare," undated. As of October 24, 2011:
http://www.hospitalcompare.hhs.gov/hospital-search.aspx