



**2015 National Impact Assessment of the
Centers for Medicare & Medicaid Services (CMS)
Quality Measures Report**

**Center for Clinical Standards and Quality,
Centers for Medicare & Medicaid Services (CMS)**

2015 National Impact Assessment of the Centers for Medicare & Medicaid Services (CMS) Quality Measures Report

Posted March 2, 2015

Prepared by: Health Services Advisory Group, Inc.

All material contained in this report is in the public domain and may be used and reprinted without special permission; citation as to source, however, is appreciated.

Suggested citation: 2015 National Impact Assessment of the Centers for Medicare & Medicaid Services (CMS) Quality Measures Report, CMS, Baltimore, Maryland, March 2, 2015. Available at:
<https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/QualityMeasurementImpactReports.html>

Table of Contents

Glossary of Acronyms/Abbreviations	iv
Executive Summary	1
Background	1
Key Findings and Actions to Consider	2
Future Directions	7
Conclusions	7
Reference List	8
Introduction.....	9
Background	9
Objectives	11
Scope of the Assessment: Programs, Time Frames, and Measures	12
Report Development	14
Data Sources and Analytic Methods.....	18
Report Structure	20
Limitations of This Report.....	22
Reference List	23
Part 1—Overview of CMS Quality Measure Use: Reach, Adoption, and Implementation ..	25
Chapter 1—CMS Measures in Relationship to the National Quality Strategy Priorities ...	26
Question on Adoption.....	26
Abstract.....	26
Background.....	27
Methods	30
Results.....	33
Conclusions.....	41
Reference List	43
Chapter 2—Measures Under Consideration: Addressing Measure Needs	44
Question on Adoption.....	44
Abstract.....	44
Background.....	45
Methods	47
Results.....	51
Conclusions.....	59
Reference List	61
Chapter 3—Physician Adoption of PQRS Measures	63
Questions on Adoption	63
Abstract.....	63
Background.....	65

Methods	66
Results.....	71
Conclusions.....	84
Reference List.....	86
Chapter 4—Measure Alignment: CMS, State, and Veterans Health Administration	
Measures	87
Questions on Adoption	87
Abstract.....	87
Background.....	88
Methods	91
Results.....	99
Conclusions.....	109
Reference List.....	111
Chapter 5—CMS Measures: Populations Reached.....	112
Questions on Reach	112
Abstract.....	112
Background.....	113
Methods	115
Results.....	121
Conclusions.....	132
Reference List.....	134
Chapter 6—Measure Use: Unintended Consequences in Hospitals, Nursing Homes, and Ambulatory Settings	136
Question on Implementation.....	136
Abstract.....	136
Background.....	138
Methods	140
Results.....	144
Conclusions.....	149
Reference List.....	151
Part 2—Overview of CMS Quality Measure Results: Effectiveness and Maintenance	157
Chapter 7—CMS Measure Trends in Performance and Disparities.....	158
Questions on Maintenance.....	158
Abstract.....	158
Background.....	160
Methods	160
Results.....	171
Conclusions.....	202
Reference List.....	205

Chapter 8—Measure Relationships: Hospital Process Measures and Patient Outcomes..	207
Questions on Effectiveness	207
Abstract	207
Background	209
Methods	211
Results	225
Conclusions	238
Reference List	240
Chapter 9—Measure Relationships: Patient-Reported Hospital Experiences and Predicted Medicare Costs.....	244
Question on Effectiveness	244
Abstract	244
Background	245
Methods	248
Results	250
Conclusions	254
Reference List	255
Chapter 10—Future Directions	258
Introduction	258
Focal Points for Future Study	258
Conclusions	260
Reference List	262

Glossary of Acronyms/Abbreviations

ACA	Patient Protection and Affordable Care Act
ACEI	Angiotensin Converting Enzyme Inhibitor
ACO	Accountable Care Organization
ADHD	Attention Deficit Hyperactivity Disorder
AHA	American Hospital Association
AHRQ	Agency for Healthcare Research and Quality
AIAN	American Indian/Alaska Native
AMI	Acute Myocardial Infarction
ARB	Angiotensin Receptor Blocker
APRN	Advanced Practice Registered Nurse
ASCQR Program	Ambulatory Surgical Center Quality Reporting Program
ASPE	Assistant Secretary for Planning and Evaluation
BH MCO	Behavioral Health Managed Care Organization
<i>C. diff.</i>	<i>Clostridium Difficile</i>
CABG	Coronary Artery Bypass Graft
CAD	Coronary Artery Disease
CAHPS	Consumer Assessment of Healthcare Providers and Systems
CART	CMS Abstraction Reporting Tool
CAUTI	Catheter Associated Urinary Tract Infection
CCW	Chronic Conditions Data Warehouse
CDC	Centers for Disease Control and Prevention
CHF	Congestive Heart Failure
CHIP	Children’s Health Insurance Program
CLABSI	Central Line Associated Bloodstream Infection
CMS	Centers for Medicare & Medicaid Services
CPT	Current Procedure Terminology
DC	District of Columbia
DO	Doctor of Osteopathy
DVT	Deep Vein Thrombosis
<i>E. coli</i>	<i>Escherichia Coli</i>
EHR	Electronic Health Record
EHR EH	Medicare and Medicaid Electronic Health Record (EHR) Incentive Program for Eligible Hospitals and Critical Access Hospitals
EHR EP	Medicare and Medicaid Electronic Health Record (EHR) Incentive Program for Eligible Professionals
EPC	Evidence-based Practice Center
eRx Incentive Program	Medicare Electronic Prescribing Incentive Program
ESRD QIP	End-Stage Renal Disease Quality Incentive Program

FASC	Federal Assessment Steering Committee
FFS	Fee-for-Service
GPRO	Group Practice Reporting Option
HAC Reduction Program	Hospital-Acquired Condition Reduction Program
HAI	Healthcare-Associated Infection
HCAHPS	Hospital Consumer Assessment of Healthcare Providers and Systems
HCC	Hierarchical Condition Category
HCPCS	Healthcare Common Procedure Coding System
HEDIS	Healthcare Effectiveness Data and Information Set
HF	Heart Failure
HH QRP	Home Health Quality Reporting Program
HHA	Home Health Agencies
HHS	U.S. Department of Health and Human Services
HMO	Health Maintenance Organization
Hospital IQR Program	Hospital Inpatient Quality Reporting Program
Hospital OQR Program	Hospital Outpatient Quality Reporting Program
Hospital VBP Program	Hospital Value-Based Purchasing Program
HQA	Hospital Quality Alliance
HQID	Hospital Quality Incentive Demonstration
HQRP	Hospice Quality Reporting Program
HRRP	Hospital Readmissions Reduction Program
HRSA	Health Resources and Services Administration
HSAG	Health Services Advisory Group
ICD-9	International Classifications of Diseases, Ninth Revision
ICU	Intensive Care Unit
IPFQR Program	Inpatient Psychiatric Facility Quality Reporting Program
IPPS	Inpatient Prospective Payment System
IRFQR Program	Inpatient Rehabilitation Facilities Quality Reporting Program
IT	Information Technology
LTCHQR Program	Long-Term Care Hospitals Quality Reporting Program
LVSD	Left Ventricular Systolic Function
MA	Medicare Advantage
MA CAHPS	Medicare Advantage Consumer Assessment of Healthcare Providers and Systems
MAP	Measure Applications Partnership
MCO	Managed Care Organization
MD	Medical Doctor
MDS	Minimum Data Set
Medicaid Adult	Core Set of Health Care Quality Measures for Adults Enrolled in Medicaid (Medicaid Adult Core Set)

Medicaid Child	Core Set of Health Care Quality Measures for Medicaid and CHIP (Child Core Set)
MEDPAC	Medicare Payment Advisory Commission
MEDPAR	Medicare Provider Analysis and Review
MRSA	Methicillin-Resistant <i>Staphylococcus Aureus</i>
MSPB	Medicare Spending Per Beneficiary
MSSP	Medicare Shared Savings Program
NCQA	National Committee for Quality Assurance
NHDR	National Healthcare Disparities Report
NHQI	Nursing Home Quality Initiative
NPI	National Provider Identifier
NPRM	Notice of Proposed Rule Making
NQF	National Quality Forum
NQS	National Quality Strategy
OASIS	Outcome and Assessment Information Set
OMB	Office of Management and Budget
ONC	Office of the National Coordinator for Health Information Technology
P4P	Pay-for-Performance
P4R	Pay-for-Reporting
Part C	Medicare Part C (Display or Star Rating)
Part D	Medicare Part D (Display or Star Rating)
PCHQR Program	Prospective Payment System-Exempt Cancer Hospitals Quality Reporting Program
PCI	Percutaneous Coronary Intervention
PCMH	Patient Centered Medical Home
PDP CAHPS	Prescription Drug Plan Consumer Assessment of Healthcare Providers and Systems
PE	Pulmonary Embolism
PI	Pacific Islander
PN	Pneumonia
POS	Point Of Service
PPO	Preferred Provider Organization
PQRS	Physician Quality Reporting System
PRA	Paperwork Reduction Act
PRO	Patient Reported Outcome
QIES	Quality Improvement System
QIN-QIO	Quality Innovation Network-Quality Improvement Organization
QIP	Quality Incentive Program
QOF	Quality and Outcomes Framework
RCT	Randomized Controlled Trial

RE-AIM	Reach, Effectiveness, Adoption, Implementation and Maintenance
SAMHSA	Substance Abuse and Mental Health Services Administration
SCIP	Surgical Care Improvement Project
SIR	Standardized Infection Ratio
SNAC	Subcommittee on Children’s Healthcare Quality Measures for Medicaid and CHIP Programs
SNP	Special Needs Plan
SSI	Surgical Site Infection
STEMI	ST-Segment Elevation Myocardial Infarction
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
TEP	Technical Expert Panel
TIA	Transient Ischemic Attack
UK	United Kingdom
U.S.	United States
UTI	Urinary Tract Infection
VA	Veterans Affairs
VBP	Value-Based Purchasing
VHA	Veterans Health Administration
VIF	Variance Inflation Factor
VTE	Venous Thromboembolism

Executive Summary

The *2015 National Impact Assessment of the Centers for Medicare & Medicaid Services (CMS) Quality Measures Report* (2015 Impact Report) is a comprehensive evaluation of the effects of the measures used in CMS quality reporting programs on achieving the goals of providing high-quality, affordable healthcare to CMS beneficiaries. The report is mandated by section 3014(b), as amended by section 10304, of the Patient Protection and Affordable Care Act (ACA), which provides that the Secretary shall, not later than March 1, 2012, and at least once every three years thereafter, conduct an assessment of the quality and efficiency impact of the use of endorsed measures described in section 1870(b)(7)(B) of the Social Security Act and make such assessment available to the public.¹ The 2012 *National Impact Assessment of Medicare Quality Measures*, encompassing trend data for eight programs from 2006 to 2010 and measures under consideration in 2012,¹ was published according to this mandate.² The 2015 Impact Report represents an assessment of 25 CMS reporting programs, using data from 2006 to 2013. The report is framed by focused research questions, organized by chapter, which were developed by a multidisciplinary Technical Expert Panel (TEP) to ascertain the impact of CMS quality measures.

Background

The gap between high-quality, evidence-based care and actual care received by patients was widely acknowledged over 10 years ago, with a corresponding interest in developing strategies to bridge this gap.^{3,4} The ACA not only mandated the assessment of the quality impact of endorsed measures, but also required the development of the National Quality Strategy (NQS). Developed through a transparent and collaborative process with input from a range of stakeholders, the NQS, first published in 2011, provides inspiration and guidance on a nationwide effort to coordinate public and private efforts to improve the quality of health and healthcare for all Americans. The NQS aligns the nation toward three shared aims of Better Care, Healthy People/Healthy Communities, and Affordable Care. These aims are



ⁱ Measures under consideration are measures that have not been finalized in previous rules and regulations for a particular CMS program and that CMS is considering for adoption through rulemaking for future implementation.

advanced through six priorities⁵ (see figure) that are mapped to the measure domains of Patient Safety (Safety), Patient and Family Engagement (Patient Engagement), Care Coordination, Clinical Process/Effectiveness (Effective Treatment), Population/Public Health (Healthy Communities), and Efficient Use of Healthcare Resources (Affordable Care). The NQS is a living and changing guide for the federal government, as well as for states and the private sector. CMS, as a key leader in transforming healthcare, and as provider to well over 100 million individuals, is a critical stakeholder in the NQS.

CMS subsequently aligned its Quality Strategy with the NQS and uses a number of “levers” to ensure the achievement of national healthcare aims and priorities for its beneficiaries. These levers include measuring performance of and publicly reporting quality measures, providing technical assistance to front-line providers and fostering quality improvement, adopting evidence-based national coverage determinations, setting clinical standards for providers that support quality improvement, and creating survey and certification processes that evaluate capacity for quality assurance and quality improvement.⁶ Critical to the success of each of these levers are quality measures. CMS quality measures and their contributions to improving the nation’s health are the focus of this 2015 Impact Report.

To aid in identifying key components of the impact assessment, the RE-AIM (Reach, Effectiveness, Adoption, Implementation, and Maintenance) framework was adapted.⁷ A Technical Expert Panel (TEP) of quality measurement leaders from across the healthcare industry applied this framework to identify and prioritize the research questions. In addition to the TEP, a Federal Assessment Steering Committee (FASC), consisting of stakeholders from CMS and other U.S. Department of Health and Human Services (HHS) agencies, convened to provide feedback. (See Appendix i-1 for a list of TEP members and Appendix i-2 for a list of FASC members.)

The 2015 Impact Report encompasses 25 CMS programs and nearly 700 quality measures from 2006 to 2013 and employs nine key research questions. Although certain analyses examined all 25 CMS programs, others examined selected measures in a few programs. Criteria used to determine the type of impact analysis, i.e., trend analysis or descriptive analysis, were as follows:

1. Trend Analysis: Measures for which sufficient performance data were available for a minimum of three consecutive years between 2006 and 2012, could be aggregated, and were publicly available.
2. Descriptive Analysis: Measures for which performance data were available for fewer than three consecutive years and in use through December 2013.

Key Findings and Actions to Consider

Measure Use

The *Measure Use* section addresses selected questions related to three RE-AIM elements: (1) *adoption* by providers and entities; (2) *reach* of measures to populations; and (3) *implementation* of measures, specifically the unintended consequences of measure implementation. Key findings and actions to consider related to these three elements include:

Key Finding: CMS achieved improved coverage and balance of quality measures addressing the six measure domains related to the NQS priorities. The findings relating to the pre-rulemaking process established by ACA section 3014 and the measures under consideration show that since 2011 CMS is making progress in establishing balanced sets of measures across its programs; however, significant gaps remain across all measure domains. The Affordable Care and Care Coordination domains are the most underrepresented.

Action to Consider 1: *Focus on Affordable Care and Care Coordination domains as high priorities for new measure development.* Novel approaches to measure design and development may be required to address these critical gaps effectively. These approaches include hybrid data sources, e.g., claims and electronic health record data, and shared accountability between providers.

Key Finding: CMS programs and measures reach a wide range of patients with high-impact conditions. The CMS Medicare quality measures reach a large majority of the top 20 high-impact Medicare conditions experienced by beneficiaries. However, measures addressing these high-impact conditions are not evenly distributed across CMS reporting programs.

Action to Consider 2: *Evaluate existing measures addressing high-impact conditions across all measure domains.* While many measures address certain high-impact conditions, e.g., cardiovascular disease, opportunities exist to evaluate the effectiveness of these measures to determine if they continue to address the CMS goals of quality and efficiency and to assess patient-centered outcomes.

Action to Consider 3: *Prioritize the high-impact conditions and develop a core set of measures across the continuum of care addressing the relevant measure domains.* This core set of measures will represent a patient-centered, cross-setting, longitudinal set of measures.

Key Finding: Less than half of the quality measures studied aligned with other state and federal programs, but additional research is needed to determine to what extent further alignment would benefit providers and patients. State Medicaid and state hospital report cards are more closely aligned than measures developed by the Veterans Health Administration (VHA). Analyses of the measures used by some state programs and the VHA showed that over half of the measures are locally developed measures.

Action to Consider 4: *Determine what degree of alignment with state and federal programs would benefit patients and providers.* Measure alignment not only reduces provider burden but also supports a multi-payer approach to transforming healthcare. However, the principle that measure specifications should be uniform is difficult to implement and should be redefined in a way that prioritizes the elements that are most important for comparability.

Action to Consider 5: *Review key sources of state and locally developed measures as a component of environmental scans conducted for new measure development efforts.* An in-depth review of the locally developed measures may suggest innovative approaches to measuring and reporting quality information that could be applied broadly and adopted

within CMS programs. This may be particularly relevant for electronic clinical quality measures developed for the VHA.

Key Finding: CMS quality measures impact patients beyond the Medicare population. Over 40 percent of the measures used in CMS quality reporting programs include individuals whose healthcare is provided by Medicaid and over 30 percent include individuals with other payer sources.

Action to Consider 6: *With new measure development, emphasize data sources, such as electronic health records and all-payer databases, to ensure measures have the widest reach across payers and populations.*

Key Finding: Among the quality measures, exclusions were varied in number and type, and provider discretion was allowed as an exclusion in over one fourth of the measures in the study. However, this study did not find patterns that systematically exclude specific populations.

Action to Consider 7: *Explore the impact of measure exclusions and the effect of measure use on the proportion of the eligible population. Measure developers can conduct the analysis during new measure development and comprehensive review to determine if the measure can effectively impact population health.*

Action to Consider 8: *Develop guidance concerning the use of measure exclusions in collaboration with the National Quality Forum (NQF) that aligns with new measure development and comprehensive review. Guidance would provide clear criteria about provider discretion exclusions and would align the use of exclusions in measure specifications for optimal usability.*

Key Finding: Literature evaluating unintended consequences of quality measurement was limited, of generally low quality, and inconclusive. A review of the literature published between 2000 and 2013 found little empirical evidence that undesired effects have occurred because of the use of quality measures.

Action to Consider 9: *Consider establishing a third-party data validation process specific to each program/setting performed by a CMS contractor or authorized vendor to ensure accurate reporting and to provide insight regarding potential unintended consequences of quality measurement. Data collected from this process could enable identification of unintended effects and may allow rapid modifications to measure specifications.*

Action to Consider 10: *Emphasize the development of balancing measures in new measure development projects. Intermediate outcome measures, e.g., glycemic control, or process measures developed as a balancing pair may mitigate the potential unintended consequences of over- or under-treatment.*

Key Finding: Provider and facility characteristics reflective of available resources (e.g., practice size, size of population served, and location) appear to be associated with increased provider participation in quality reporting programs and higher performance on quality measures.

Nursing homes are an exception to this finding as urban, non-government, and large nursing home facilities had fewer high-performing measures than their counterparts did.

Action to Consider 11: *Consider whether and to what extent the Quality Innovation Network-Quality Improvement Organizations (QIN-QIOs), as part of the 11th Scope of Work, can be directed to provide more focused technical assistance to smaller hospitals and physician practices and larger nursing homes.*

Measure Results

The *Measure Results* section of the report addresses selected questions related to the RE-AIM elements: (1) *maintenance* of measure results, including trends and disparities in measure performance, and (2) *effectiveness* of measures, including changes in health outcomes and costs related to measure performance. For the trend analysis, a limited number of measures (n=119) and reporting programs (n=7) were analyzedⁱⁱ. Key findings and actions to consider related to these two RE-AIM elements include:

Key Finding: Ninety-five percent of 119 publicly reported measure rates across seven quality reporting programs showed improvement during the study period (2006–2012). Measures that address clinical guidelines for patient care (process measures) were most likely to be high performing (i.e., measure rates exceeding 90 percent in the three final years for which data were available), suggesting process measures are more sensitive to provider quality improvement efforts than outcome measures. Process measures may have a limited lifespan, since performance benchmarks are more rapidly achieved. Few measures that addressed clinical outcomes were high performing, though most demonstrated consistent improvements over the study period.

Action to Consider 12: *Reserve the development of process measures to those processes of care that link directly to patient outcomes and in which significant variation in performance exists across providers. Conversely retire existing process measures that do not meet these criteria. Although providers' performance improves more slowly for outcome measures than process measures, the latter can provide important real-time information on the progress of quality improvement efforts to both CMS and providers. This will allow CMS and providers to make contemporaneous adjustments to program policies and quality intervention.*

Key Finding: Approximately 35 percent of the 119 measures were classified as high performing, i.e., measure rates exceeding 90 percent in each of the most recent three years for which data were available. Further improvement on these measures may provide marginal returns in terms of impact on patient outcomes. CMS has started to develop objective criteria related to measure retirement, specific to individual reporting programs; however, an approach could be considered related to measure retirement that is sensitive to any identified disparities.

ⁱⁱ Measures included in this analysis had a minimum of three consecutive years of data collected between 2006 and 2012, could be aggregated, and were publicly reported.

Action to Consider 13: *Develop clinically valid performance thresholds for high-performing measures that take into account any identified disparities, above which individual provider rates would be publicly reported as “exceeding performance standards.”* Once measure rates exceed established performance thresholds, providers can focus resources on quality measures that do not meet performance thresholds.

Action to Consider 14: *Develop standardized criteria in collaboration with NQF to retire quality measures.* The criteria should include an assessment of national provider performance that accounts for overall performance means, distribution of measure rates, and disparity analysis. These criteria will objectively and uniformly inform the decision to retire measures from a reporting program. Measure developers could apply the criteria during the NQF comprehensive review, which is currently on a three-year cycle.

Key Finding: Widespread race and ethnicity disparities that existed at the beginning of the study period in 2006 were much less pronounced in 2012; however, disparities persist across select programs, settings, and demographic groups. While lack of comparability in how race and ethnicity are assessed across settings limits generalizability, measure rates for Hispanics, Blacks, and Asians improved the most, and measure rates for American Indian/Alaska Natives and Native Hawaiian/Pacific Islanders improved the least.

Action to Consider 15: *Standardize race and ethnicity data collection across CMS quality reporting programs to facilitate disparity analysis.* Implementation of standardized data elements will increase generalizability of analyses across reporting programs.

Action to Consider 16: *Promote transparency and decrease the disparities in measure performance by publicly reporting and systematically monitoring the measure rates by race and ethnicity for all publicly reported measures.* These measures results can be compared within peer groups of providers serving disproportionately large minority and underserved populations.

Key Finding: A positive relationship was found between performance on a limited number of CMS measures and positive patient outcomes; however, the effects were variable, and a small number of process measures were estimated to have an impact on the health of the eligible Medicare population. CMS process measures related to heart and surgical care were linked to improved patient outcomes. From 2006 to 2012, between 7,000 and 10,000 lives were saved through improved performance on inpatient hospital heart failure process measures, and between 4,000 and 7,000 infections were averted through improvement in performance on inpatient hospital surgical process measures. Lower-than-expected Medicare costs had a mild but persistent relationship to greater patient satisfaction with hospital inpatient experiences using 2012 data.

Action to Consider 17: *Develop more outcome measures, including patient-reported outcomes, since process measures are not uniformly achieving better health outcomes.* Prioritization of outcome measures and the development of standard methodologies, e.g., risk adjustment, may facilitate outcome measure development.

Action to Consider 18: *Quantify and monitor process-outcome linkages during measure implementation.* Within the CMS reporting programs, measure developers should quantify the strength of the relationship between the process of care and the corresponding health outcome and monitor this linkage during implementation. The measure developers could quantify this linkage as part of the annual and three-year comprehensive review required by NQF.

Action to Consider 19: *Conduct a qualitative and quantitative study to understand the characteristics of an organization or provider that are associated with high performance on outcome and cost measures.* Explore what factors may account for positive patient outcomes in facilities with lower-than-expected costs; for example, improved care coordination may improve patient experience and reduce readmissions, while contributing to lower costs. This study should include an analysis of disparities factors, i.e., differences by race/ethnicity, poverty, disability status, etc. and their relationship to outcomes.

Future Directions

The 2018 Impact Report will highlight progress on the CMS Quality Strategy, while endeavoring to provide new insights for making informed measure and program-specific decisions. Important focal points for the 2018 Impact Report include:

- ◆ A greater number of program measures and analyses—data that were not accessible for recently introduced programs will be available.
- ◆ Analysis of data at the patient level—patient-level data will provide information about beneficiary characteristics that affect measure performance that may be masked by facility-level data.
- ◆ A qualitative study on the impact of quality measures at the provider level—a national provider survey will provide an in-depth examination of the impact CMS measures are having on providers.
- ◆ A more explicit treatment of health disparities to support the foundational principle of the CMS Quality Strategy—disparities analyses will be conducted for each applicable research question.

Conclusions

Quality measurement is a key lever that CMS uses to transform delivery of healthcare. The results of this impact assessment illustrate how CMS quality measures directly support the CMS Quality Strategy goals and substantially and comprehensively contribute to national healthcare aims. This 2015 Impact Report will help CMS, providers, private payers, and communities to understand which measures have worked well and which have had less impact on quality. This will inform ongoing refinement of CMS quality measurement strategies. Everyone receiving healthcare in the nation is likely to benefit from CMS programs and initiatives, as healthcare professionals engage in delivery system reform to achieve better care for patients, better health for the U.S. population and lower costs through quality improvement.

Reference List

- (1) The Patient Protection and Affordable Care Act, Pub. L. No. 111–148, § 3014(b) 124 Stat. 119, 385–87 (2010) (as amended at 42 U.S.C. § 1395aaa-1).
- (2) Centers for Medicare & Medicaid Services. *National Impact Assessment of Medicare Quality Measures*. Baltimore, MD: Centers for Medicare & Medicaid Services; March 2012.
- (3) Institute of Medicine (US). *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington DC: National Academies Press; 2001.
- (4) McGlynn EA, Asch SM, Adams J, et al. The quality of health care delivered to adults in the United States. *N Engl J Med*. 2003;348(26):2635-264.
- (5) U.S. Department of Health and Human Services. *Report to Congress: National Strategy for Quality Improvement in Health Care*. Washington DC: U.S. Dept of Health and Human Services; 2011.
- (6) Centers for Medicare & Medicaid Services. *CMS Quality Strategy - 2013 - Beyond*. Baltimore, MD: Centers for Medicare & Medicaid Services; November 18, 2013.
- (7) Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *Am J Public Health*. 1999;89(9):1322-1327.

Introduction

Background

The Patient Protection and Affordable Care Act (ACA), section 3014(b) as amended by section 10304, states that not later than March 1, 2012, and at least once every three years thereafter, the Secretary of Health and Human Services (HHS) shall conduct an assessment of the quality and efficiency impact of the use of endorsed measures described in section 1890(b)(7)(B) of the Social Security Act and make such assessment available to the public.¹ According to this mandate, the *National Impact Assessment of Medicare Quality Measures* (2012 Impact Report) was published in March 2012; it includes the measure data from 2006 to 2010 for eight programs and the measures under considerationⁱⁱⁱ in 2012.² This *2015 National Impact Assessment of the Centers for Medicare & Medicaid Services (CMS) Quality Measures Report* (2015 Impact Report) presents the results of the second impact assessment of the quality measures used by CMS.

Driving quality improvement is a core function of CMS. The CMS vision is to have a high quality healthcare system that ensures better care, access to coverage, and improved health.⁴ To accomplish this vision, CMS has aligned the goals of its Quality Strategy (published in November 2013) with the six priorities of the National Quality Strategy (NQS).^{3,5} CMS has mapped the NQS priorities to six measurement domains: Patient Safety (Safety); Patient and Family Engagement (Patient Engagement); Care Coordination (Care Coordination); Clinical Process/Effectiveness (Effective Treatment); Population/Public Health (Healthy Communities); and Efficient Use of Healthcare Resources (Affordable Care). The domains are used throughout this report to categorize measures and the subsequent results of the analyses conducted. Table i-1 shows the alignment of the NQS priorities, the CMS Quality Strategy goals and objectives, and the measure domains.

As CMS operationalizes the goals of the CMS Quality Strategy, it is guided by four foundational principles: (1) eliminate racial and ethnic disparities, (2) strengthen infrastructure and data systems, (3) enable local innovations, and (4) foster learning organizations. The foundational principles are incorporated into the plans for implementing specific initiatives and activities, including measuring and public reporting of quality performance by providers. Quality measurement is a key driver to improve care in CMS quality initiatives. CMS developed its first set of standardized quality measures for managed care plans in 1997 and began publicly reporting the results in 1999.⁶ Since then, CMS has developed and adopted many additional quality measures to reflect the evolving needs and priorities of the country's healthcare system. The measures used in the CMS programs increase transparency and serve as the foundation for CMS quality initiatives that improve the delivery and quality of care received by patients.

ⁱⁱⁱ Measures under consideration are quality and efficiency measures being considered for use in Medicare programs. U.S. Department of Health and Human Services (HHS) is required to publish these measures annually.

Table i-1: National Quality Strategy and CMS Quality Strategy

National Quality Strategy Priorities	CMS Quality Strategy Goals and Objectives	Measure Domains (Abbreviated)
1. Making care safer by reducing the harm caused in the delivery of care	Goal 1: Make care safer by reducing harm caused in the delivery of care <ul style="list-style-type: none"> ◆ Improve support for a culture of safety ◆ Reduce inappropriate and unnecessary care ◆ Prevent or minimize harm in all settings 	Patient Safety (Safety)
2. Ensuring that each person and family are engaged as partners in their care	Goal 2: Strengthen person and family engagement as partners in their care <ul style="list-style-type: none"> ◆ Ensure all care delivery incorporates patient and caregiver preferences ◆ Improve experience of care for patients, caregivers, and families ◆ Promote patient self-management 	Patient and Family Engagement (Patient Engagement)
3. Promoting effective communication and coordination of care	Goal 3: Promote effective communication and coordination of care <ul style="list-style-type: none"> ◆ Reduce admissions and readmissions ◆ Embed best practices to manage transitions to all practice settings ◆ Enable effective healthcare system navigation 	Care Coordination (Care Coordination)
4. Promoting the most effective prevention and treatment practices for the leading causes of mortality, starting with cardiovascular disease	Goal 4: Promote effective prevention and treatment of chronic disease <ul style="list-style-type: none"> ◆ Increase appropriate use of screening and prevention services ◆ Strengthen interventions to prevent heart attacks and strokes ◆ Improve quality of care for patients with multiple chronic conditions ◆ Improve behavioral health access and quality care ◆ Improve perinatal outcomes 	Clinical Process/Effectiveness (Effective Treatment)
5. Working with communities to promote wide use of best practices to enable healthy living	Goal 5: Work with communities to promote best practices of healthy living <ul style="list-style-type: none"> ◆ Partner with and support federal, state, and local public health improvement efforts ◆ Improve access within communities to best practices of healthy living ◆ Promote evidence-based community interventions to prevent and treat chronic disease ◆ Increase use of community-based social services support 	Population/Public Health (Healthy Communities)
6. Making quality care affordable for individuals, families, employers, and governments by developing and spreading new healthcare delivery models (Affordable Care)	Goal 6: Make care affordable <ul style="list-style-type: none"> ◆ Develop and implement payment systems that reward value over volume ◆ Use cost analysis data to inform payment policies 	Efficient Use of Healthcare Resources (Affordable Care)

The quality measures themselves must continuously undergo review and improvement. As Chassin and colleagues observed, “The process of improving our system of high-stakes quality measurement requires perpetual vigilance ... A vital part of this ... will be a formal process of assessing experience with the measures and using that information to improve the development of measures and decisions regarding deployment.”⁷ Toward this end, the Institute of Medicine (IOM) cites the “continuously learning healthcare system” as a mechanism to change and improve its actions and outputs over time. IOM proposes a cycle of evidence-based care, which contributes to understanding how to best improve care and avoid past mistakes, ensuring that care provided by clinicians for patients and communities results in continuously improving care. IOM notes that this process requires an interdisciplinary approach supported by commitment from leadership, incentives for growth, and a culture of quality.⁸ These perspectives of continuous learning and vigilance emphasize the importance of regularly examining the impact of quality measures.

The 2015 Impact Report is distinct from program evaluation in that the focus is on the measures themselves and the role they play in leveraging results and reflecting progress toward CMS goals. This report builds on knowledge of CMS measure performance trends provided in the 2012 Impact Report and introduces a number of in-depth analyses. The trend data and analyses examine multiple dimensions of CMS quality measure use and results. These analyses do not replace or duplicate program-specific assessments, nor do they replace the individual measure analyses that are important to ongoing measure maintenance. Rather, they are intended to help the federal government understand the impact of its investments in quality measurement as a key driver of improvements in the delivery and quality of care received by patients.

Objectives

Conducted over three years using available datasets spanning 2006 to 2013, the 2015 Impact Report contributes to the crosscutting evaluation of CMS quality measures through three objectives:

Objective 1—Examine the Extent of Current Measure Use

CMS collects data from hundreds of measures that assess the quality of care delivered in a range of settings, including hospitals, nursing homes, home health agencies, and clinician offices.⁹ Other entities throughout the nation’s healthcare system also have developed and adopted quality measures. The RAND Corporation recently estimated that more than 580 health-related organizations now disseminate quality measures, including many CMS measures.¹⁰ Analyses of measure use included in this report examine the following issues:

- ◆ The extent to which CMS quality measures address NQS priorities.
- ◆ How the pre-rulemaking process of evaluating measures that CMS is considering for implementation across CMS programs has changed the proportion of CMS measures addressing NQS priorities over time.
- ◆ Whether and why there may be differences in adoption of measures by physicians in the Physician Quality Reporting System (PQRS).

- ◆ Whether measures used by state agencies and the Veterans Health Administration (VHA) are aligned with the measures used by CMS.
- ◆ The degree to which populations are reached by the quality measures in Medicare programs and whether populations are underrepresented by CMS measures.
- ◆ The extent to which there is published evidence regarding the unintended consequences related to the implementation and use of quality measures and the effect of these unintended consequences on the healthcare system.

Objective 2—Analyze Measure Results

CMS quality improvement initiatives and incentive programs have generated sizable datasets, many of which span numerous years. This assessment of the impact of CMS quality measures applies these datasets to examine measure results related to the following dimensions:

- ◆ Trends in performance and disparities across settings.
- ◆ The relationship between process-of-care measures and health outcomes.
- ◆ The relationship between patient-reported experiences of care and predicted Medicare costs.

Objective 3—Lay the Foundation for New Frontiers

In an essay for the *Journal of the American Medical Association*, Dr. Patrick Conway, Deputy Administrator for Innovation & Quality and CMS Chief Medical Officer and colleagues wrote: “Meaningful quality measures increasingly need to transition from setting-specific, narrow snapshots ... to assessments that are broad based, meaningful, and patient centered in the continuum of time in which care is delivered.”¹¹ CMS continually works with policymakers, practitioners, and key stakeholders to streamline measures across the healthcare continuum. The 2015 Impact Report examines results that will provide guidance for transitioning to the next generation of quality measures. By examining the reach and performance of the measures across settings and over years, the 2015 Impact Report provides insights for the development of new concepts and topics to be addressed in the 2018 Impact Report.

Scope of the Assessment: Programs, Time Frames, and Measures

The 2015 Impact Report focuses on the quality measures identified in section 3014 of the ACA, and examines measure results spanning up to seven years within and across CMS quality reporting and value-based purchasing programs. Other National Quality Forum (NQF)-endorsed and non-endorsed measures used for quality reporting across 25 CMS programs and initiatives have also been incorporated. Furthermore, one chapter of this 2015 Impact Report addresses the measures being considered by CMS for future use (measures under consideration) that are part of the pre-rulemaking process mandated by section 3014 of the ACA. CMS submits these measures under consideration annually for use in a variety of CMS quality reporting and value-based purchasing programs and is required to do so for any Medicare measures described in section 1890(b)(7)(B) of the Social Security Act. The NQF-convened Measure Applications Partnership (MAP) provides stakeholder feedback on these measures.

The 25 quality measurement programs and initiatives (referred to as “programs” throughout this 2015 Impact Report) are organized in Table i-2 according to three healthcare delivery settings: Hospital, Ambulatory, and Post-Acute. Table i-2 also notes the abbreviations used throughout this report to identify the programs.

Table i-2: Programs Included in the 2015 Impact Report

Setting	Program	Abbreviation
 Hospital	Hospital Inpatient Quality Reporting Program	Hospital IQR Program
	Hospital Value-Based Purchasing Program	Hospital VBP Program
	Hospital Readmissions Reduction Program	HRRP
	Hospital-Acquired Condition Reduction Program	HAC Reduction Program
	Medicare and Medicaid Electronic Health Record (EHR) Incentive Program for Eligible Hospitals and Critical Access Hospitals	EHR EH
	Hospital Outpatient Quality Reporting Program	Hospital OQR Program
	Ambulatory Surgical Center Quality Reporting Program	ASCQR Program
	Inpatient Psychiatric Facility Quality Reporting Program	IPFQR Program
	Prospective Payment System-Exempt Cancer Hospitals Quality Reporting Program	PCHQR Program
 Ambulatory	Physician Quality Reporting System	PQRS
	Medicare Electronic Prescribing Incentive Program	eRx Incentive Program
	Physician Feedback Program	Physician Feedback Program
	Medicare and Medicaid Electronic Health Record (EHR) Incentive Program for Eligible Professionals	EHR EP
	Medicare Shared Savings Program	MSSP
	Physician Compare	Physician Compare
	Medicare Part C (Display or Star Ratings)	Part C
	Medicare Part D (Display or Star Ratings)	Part D
	Core Set of Children’s Health Care Quality Measures for Medicaid and CHIP (Child Core Set)	Medicaid Child
	Core Set of Health Care Quality Measures for Adults Enrolled in Medicaid (Medicaid Adult Core Set)	Medicaid Adult
 Post-Acute	Nursing Home Quality Initiative	NHQI
	Home Health Quality Reporting Program	HH QRP
	End-Stage Renal Disease Quality Incentive Program	ESRD QIP
	Hospice Quality Reporting Program	HQRP
	Inpatient Rehabilitation Facilities Quality Reporting Program	IRFQR Program
	Long-Term Care Hospitals Quality Reporting Program	LTCHQR Program

Report Development

Health Services Advisory Group and the RAND Corporation, its subcontractor, developed this report over three years. This research team consisted of highly specialized individuals from multiple disciplines with expertise in healthcare, quality measures, health services research, and healthcare policy. This section describes the activities that are related to the report oversight and review, assessment framework, logic model development, and research question development. These activities ensured that the report incorporates key questions relevant to the impact of CMS quality measures and that major stakeholders had the opportunity to provide input during all stages of report development.

Report Oversight and Review

Technical Expert Panel

The research team convened a Technical Expert Panel (TEP) to provide input on the development of an analytic plan and assessment activities. The TEP comprises 18 national healthcare policy experts, researchers, and practitioners with extensive knowledge of and experience with CMS quality measures. George Isham, MD, MS, who is also co-chair of the NQF-convened Measure Applications Partnership Coordinating Committee, chaired the TEP. The TEP first met in Baltimore, Maryland, in March 2012 and continued to provide expert input throughout the development and finalization of this report, meeting six times from 2013 to 2014. Appendix i-1 lists the TEP members.

Federal Assessment Steering Committee

The Federal Assessment Steering Committee (FASC) was formed to participate in the planning and oversight of the 2015 Impact Report and to advise CMS on relevant policy and/or measurement issues. The FASC comprises 13 members from federal agencies, including CMS, the Agency for Healthcare Research and Quality (AHRQ), the Health Resources and Services Administration (HRSA), the Office of the Assistant Secretary for Planning and Evaluation (ASPE), the Office of the National Coordinator for Health Information Technology (ONC), the Centers for Disease Control and Prevention (CDC), and the Substance Abuse and Mental Health Services Administration (SAMHSA). The FASC met six times between February and July 2014 to provide input on study results. A roster of FASC members is in Appendix i-2.

Comments from both the TEP and the FASC were carefully considered and systematically incorporated into the report. Upon completion of the draft study analyses, the research team presented the findings to the TEP and the FASC for review and feedback.

Assessment Framework

RE-AIM

With the help of the TEP, the research team reviewed a variety of impact assessment frameworks.¹²⁻¹⁵ The RE-AIM framework was selected as the most appropriate tool to ensure

research questions addressed key elements of the impact of quality measures. The RE-AIM framework was originally designed to evaluate the public health impact of health promotion interventions¹² and has been used in over 150 published studies.¹⁶ Since quality measurement can be viewed as a type of intervention to improve quality and population health, the research team and the TEP adapted the RE-AIM framework to determine the role of CMS measures in transforming healthcare. The adapted definitions of the RE-AIM elements are as follows:

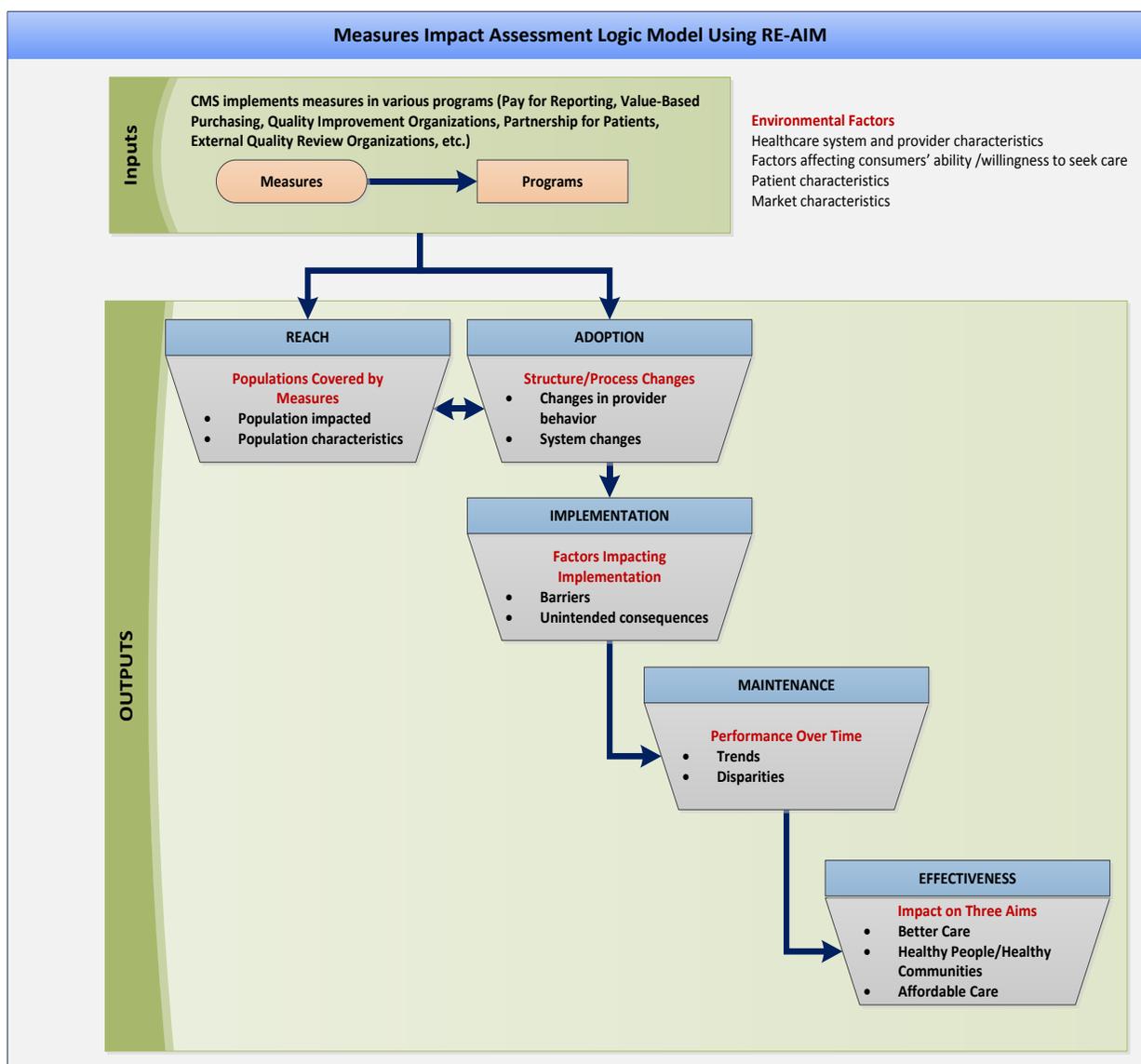
- ◆ **Reach** refers to the extent to which a set of measures represents a given population. Reach is measured by the number, proportion, and representativeness of beneficiaries (intended or actual) captured by a given measure.
- ◆ **Effectiveness** refers to changes in outcomes, quality of care, and healthcare costs that are associated with measure use. Of interest are changes at both the patient level and the population level.
- ◆ **Adoption** refers to the extent to which providers, whether individuals or facilities, have adopted quality measures appropriate to their practice and is measured by the number, proportion, and representativeness of providers who are using CMS quality measures and reporting measure data to CMS.
- ◆ **Implementation** refers to the factors affecting impact that are associated with the application of a given measure or measure set within a program, e.g., barriers to reporting of a measure by provider.
- ◆ **Maintenance** refers to the extent to which measure use has become institutionalized or incorporated into routine organizational practices and policies. Maintenance includes measure performance over time, particularly evaluation of disparities between populations.

Logic Model Development

Using the modified RE-AIM framework, the research team, in collaboration with the TEP, developed the logic model presented in Figure i-1. The logic model illustrates the hypothesized relationships between the research questions within each RE-AIM element and the three aims of Better Care, Healthy People/Healthy Communities, and Affordable Care and includes the following:

- ◆ Inputs, e.g., measures and the CMS reporting programs that use them, including public reporting, value-based purchasing, and quality improvement programs.
- ◆ Environmental factors, e.g., the healthcare delivery system or organizational characteristics.
- ◆ Outputs, e.g., reach of measures, adoption of measures, changes in provider behavior, barriers to measure implementation, measure trends, and effects on healthcare costs.

Figure i-1: Measure Impact Assessment Logic Model Using RE-AIM



Research Question Development

Using the revised RE-AIM framework and the logic model, the TEP identified over 40 candidate research questions during an initial meeting in March 2012. The TEP members then participated in a modified Delphi method process¹⁷ to arrive at a select list of research questions. The modified Delphi method involved rating attributes of each research question against two criteria: (1) Relevance, defined as the degree to which a research question addresses the CMS goal of understanding the impact of its measurement programs on the three aims, and (2) Importance, defined as how critical a research question is to identifying the impact of CMS measures. Ratings were performed on a 1–9 scale, defined as follows: *Relevance* (1=Not relevant at all; 5=Relevance unclear; 9=Definitely relevant), *Impact* (1=Not important at all; 5=Importance unclear; 9=Definitely important).

The TEP met to review rating results in June 2012 and discussed in detail research questions for which there was no clear agreement. Three days later, the TEP members independently re-rated the questions. The highest rated questions were subsequently approved by CMS for inclusion in the 2015 Impact Report. Although no single research question can comprehensively assess the impact of CMS measures, the final set of research questions address all five dimensions of the RE-AIM framework. The research questions were linked to the benefits of improved healthcare for patients and families (Table i-3).

Table i-3: 2015 Impact Report Chapter Topics and Impact on Patients and Families

Chapter	Impact on Patients and Families
Chapter 1 CMS Measures in Relationship to the National Quality Strategy Priorities	Chapter 1 examines how CMS is addressing the NQS priorities in the measures used in its programs. The aims and priorities were established to improve the quality of the healthcare provided to individual patients, thereby leading to better population health and better affordability of care.
Chapter 2 Measures Under Consideration: Addressing Measure Needs	Chapter 2 examines how CMS is filling gaps in the measures used in its Medicare programs. By filling these gaps, CMS will enhance its ability to measure additional aspects of the quality of care provided to patients and families.
Chapter 3 Physician Adoption of PQRS Measures	Chapter 3 identifies the characteristics of physicians participating in a voluntary quality reporting program. Understanding these characteristics will enable CMS to better design quality initiatives that will result in better care for patients and their families.
Chapter 4 Measure Alignment: CMS, State, and Veterans Health Administration (VHA) Measures	Chapter 4 determines the extent of measure alignment between CMS, state programs, and the VHA. Alignment of quality measures will bring a synergistic effect in driving improved care for patients across all communities in the country.
Chapter 5 CMS Measures: Populations Reached	Chapter 5 examines the exclusions used in quality measure specifications. This information provides insights on whether balance has been achieved between creating standardized exclusions to allow for specificity in quality measurement and the need to allow for provider and patient decision making that is appropriate for the patient's current care needs.
Chapter 6 Measure Use: Unintended Consequences in Hospitals, Nursing Homes, and Ambulatory Settings	Chapter 6 reports the results of a systematic review of the literature to assess whether quality measurement has undesired effects on the care provided to patients and families. These results can be used to address unintended consequences proactively that might cause harm to patients.
Chapter 7 CMS Measure Trends in Performance and Disparities	Chapter 7 examines the trends in quality measurement results. CMS can assess whether its quality measurement programs are achieving the results of improving care for the beneficiaries. Through the identification of disparities, appropriate interventions can be developed to ensure that all patients receive the same level of high-quality care.
Chapter 8 Measure Relationships: Hospital Process Measures and Patient Outcomes	Chapter 8 demonstrates the positive effect that quality measures may have had in saving lives and reducing hospitalizations for Medicare patients. CMS can identify and develop measures that impact patient outcomes.
Chapter 9 Measure Relationships: Patient-Reported Hospital Experiences and Predicted Medicare Costs	Chapter 9 focuses on quality as measured from the patient's point of view. Analyses demonstrate whether or not additional utilization of services surrounding hospitalization events is associated with positive or negative patient perceptions of care received.

Data Sources and Analytic Methods

The 2015 Impact Report encompasses 25 CMS Medicare and Medicaid programs and nearly 700 quality measures from 2006 to 2013 and employs nine key research questions. Although certain analyses examined all 25 CMS programs, others examined selected measures in a few programs. Criteria used to determine the type of impact analysis, i.e., trend analysis or descriptive analysis, were as follows:

1. **Trend Analysis:** Measures for which sufficient performance data were available for a minimum of three consecutive years between 2006 and 2012, could be aggregated, and were publicly available.
2. **Descriptive Analysis:** Measures for which performance data were available for fewer than three consecutive years and in use through December 2013. For example, programs that have recently been implemented, such as HQR, LTCHQR, and IRFQR, do not have adequate data.

A hyperlink to a comprehensive list of measures included in each analysis (indicated by chapter number) can be found in Appendix i-4. The list includes the NQF endorsement status, NQF number if endorsed, and both the measure title used by the CMS program and the measure title used by NQF. Appendix i-5 illustrates the total number of measures within settings and by program as well as the number of measures aligned^{iv} across multiple programs as of December 31, 2013. Appendix i-5 is a key reference in understanding the relationships between the performance of programs and their individual measure sets.

The analytic methods are multifaceted, including quantitative analyses, qualitative analyses, or mixed methods, using provider-level and patient-level data, and capture a range of perspectives. Though the methods for certain questions feature conventional approaches, e.g., descriptive analyses or multiple regression analyses, the methods for other questions use statistical strategies in novel ways to extract policy-relevant results. Data sources include CMS, CMS measure development contractors, CMS websites, interviews with healthcare providers, and other external databases, e.g., the American Hospital Association provider database that reports on quality measures and characteristics of patients and providers.

The analyses evaluate measure impact at a single point in time and longitudinally, depending on the research question. Table i-4 provides a brief overview of the analytic methods used for each research question. Further details about the data sources and time frames are found in each chapter.

^{iv} Although a measure may be identified as aligned in Appendix i-5, the measure specifications and/or reporting requirements may vary by program.

Table i-4: Overview of the 2015 Impact Report Research Questions and Analytic Methods

Research Question	Analytic Method
<i>Measure Use Research Questions</i>	
<p>Chapter 1—CMS Measures in Relationship to the National Quality Strategy (NQS) Priorities</p> <p>What is the progress on addressing the NQS priorities as assessed by the CMS quality measures?</p>	<p>Descriptive and comparative analyses of the NQS priorities addressed by measures used in CMS quality measurement programs as well as measures under consideration for future use.</p>
<p>Chapter 2—Measures Under: Addressing Measure Needs</p> <p>How are the gaps in Medicare measures being addressed through the pre-rulemaking process?</p>	<p>Descriptive analysis of the number and types of measures under consideration submitted for future use.</p>
<p>Chapter 3—Physician Adoption of PQRS Measures</p> <p>Which physician and patient characteristics are associated with physician participation in the voluntary PQRS?</p> <p>Which physician and patient characteristics are associated with the types of measures PQRS participants chose to report?</p>	<p>Quantitative analysis of physician characteristics associated with the decision to participate in PQRS.</p>
<p>Chapter 4—Measure Alignment: CMS, State, and Veterans Health Administration (VHA) Measures</p> <p>What measures do the states and the VHA adopt?</p> <p>How are the measures used by the states and VHA aligned with measures used by CMS?</p>	<p>Descriptive and comparative analyses of more than 1,000 measures using publicly available data from state health department entities, state Medicaid programs, and the VHA.</p>
<p>Chapter 5—CMS Measures: Populations Reached</p> <p>What populations are reached by the quality measures used in Medicare programs?</p> <p>Are there populations that are systematically excluded from the measures?</p>	<p>Descriptive and comparative analyses of denominator inclusions and exclusions for CMS measures in 25 measure programs, including currently implemented and finalized measures.</p>
<p>Chapter 6—Measure Use: Unintended Consequences in Hospitals, Nursing Homes, and Ambulatory Settings</p> <p>Has the implementation of quality measures been associated with unintended consequences?</p>	<p>Systematic review of the literature on unintended consequences of quality measure use.</p>
<i>Measure Results Research Questions</i>	
<p>Chapter 7—CMS Measure Trends in Performance and Disparities</p> <p>To what extent did providers' performance on the quality measures improve over time?</p> <p>What are the disparities in measure rates for age, sex, race, and ethnicity, and how did these disparities change over time?</p>	<p>Quantitative analysis: Trend analysis using effect size metrics to examine how quality measures in seven CMS measure programs perform over time and how CMS measures that address each NQS priority perform over time. Descriptive comparisons of effect size results to assess how measure performance over time varies among provider characteristics (e.g., bed size, urbanicity,^v teaching status) and population-level demographics (e.g., race, ethnicity, gender). Disparity analysis was performed using regression analysis to determine if the differences in quality measure rates across patient demographic groups were diminishing over time.</p>

^v *Urbanicity* is a term used to define the degree to which a geographical unit is urban.

Table i-4: Overview of the 2015 Impact Report Research Questions and Analytic Methods

Research Question	Analytic Method
<p>Chapter 8—Measure: Hospital Process Measures and Patient Outcomes</p> <p>Are changes in performance rates for clinical process-of-care measures associated with changes in patient outcomes?</p> <p>Are changes in performance rates for clinical process-of-care measures associated with changes in health outcomes of the target population?</p>	<p>Mixed methods: Qualitative examination to identify the process measures used in nine CMS measure reporting programs that are most likely to be related to one or more outcome measures used in these programs. Follow-up quantitative analyses to determine whether there is a statistically significant relationship between each identified process measure and its paired outcome measure after controlling for key patient/provider characteristics.</p>
<p>Chapter 9—Measure Relationships: Patient-Reported Hospital Experiences and Predicted Medicare Costs</p> <p>Is there a relationship between hospital-level patient-reported experience of care and risk-adjusted Medicare spending for hospitalization and 30-day post-hospital care?</p>	<p>Quantitative analysis: Regression analysis that examines whether a composite measure of Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS)^{vi} is associated with Medicare Spending Per Beneficiary (MSPB)^{vii}, after controlling for key hospital and patient characteristics.</p>

Report Structure

Chapter Organization

The 2015 Impact Report is divided into two parts: Measure Use and Measure Results. Measure Use addresses research questions related to three RE-AIM elements: (1) *adoption* by providers and entities, (2) *reach* of measures to populations, and 3) *implementation* of measures, specifically the unintended consequences of measure implementation. The Measure Results section addresses research questions related to the RE-AIM elements: (1) *maintenance of measure results*, including trends and disparities in measure performance and (2) *effectiveness* of measures, including changes in health outcomes and costs related to measure performance.

Each chapter includes the RE-AIM dimension, the research question(s) being addressed, the analytic results, and the policy and measurement implications of those results. Methods for each analysis include concise descriptions with enough detail to evaluate the strength of the analytic plan. Technical descriptions of data, methods, and analyses are presented in the appendices.

Figure i-2 illustrates whether the research questions are in the Measure Use or the Measure Results section and color-codes the questions by the RE-AIM elements. The Measure Use subset of questions (chapters 1–6) includes research questions that use largely descriptive approaches to examine measure data within and across numerous CMS quality initiatives or incentive programs. In line with the objectives of the 2015 Impact Report, these questions examine the extent to which CMS program measures reflect national health priorities, engage healthcare providers, reach patient populations, and function as intended. This broad perspective allows for reflection on the evolution of the CMS quality measurement programs, indicates progress toward addressing NQS aims and priorities, and anticipates future directions for

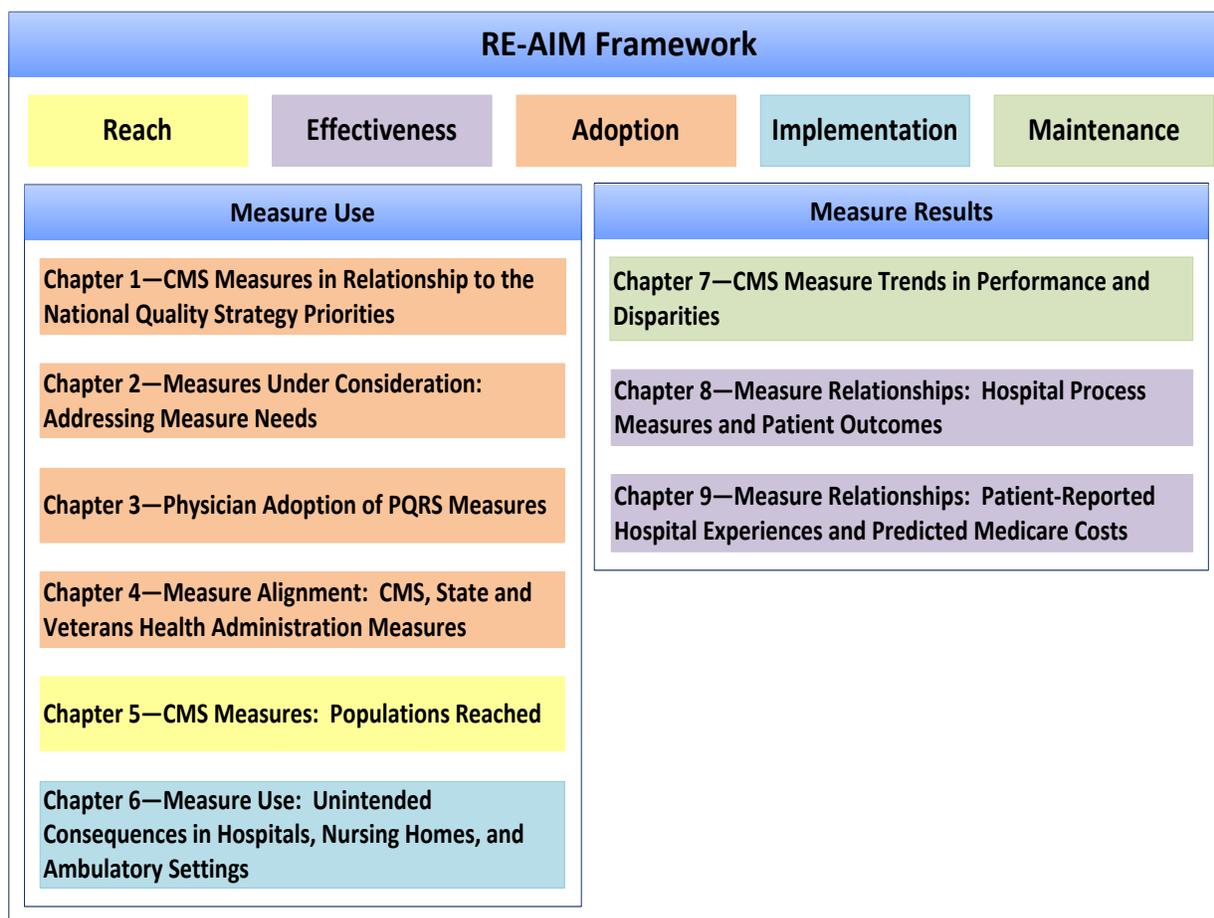
^{vi} NQF-endorsed title: *HCAHPS* (NQF #0166).

^{vii} NQF-endorsed title: *Payment-Standardized Medicare Spending Per Beneficiary (MSPB)* (NQF #2158).

enhancing measure use. These questions also provide the groundwork for the next set of analyses.

The second subset of questions (Measure Results, chapters 7–9) delves into selected CMS quality initiatives and incentive programs. The intent of this section is to evaluate performance of quality measures as well as to identify detailed patterns, gaps, and insights that refine the development and implementation of quality measures. The findings apply directly to the investigated quality measure programs, but also yield important cross-cutting information that applies to all measure-driven quality initiatives and incentive programs.

Figure i-2: Chapter Titles and Research Questions



Future Directions

This report concludes with a reflection on key findings learned through the process of conducting this assessment. Potential topics to be included in the 2018 Impact Report are provided.

Appendices

Appendices of this report, organized by each chapter, provide details related to the process of conducting the impact assessment, including participant rosters for the TEP and FASC, measures analyzed, information and data tables, and detailed study methodology.

Limitations of This Report

The research team encountered a number of challenges that affected the type and extent of analyses performed. As with many comprehensive assessments, integrating and synthesizing data across diverse programs at varying stages of implementation can lead to limitations in the ability to align measures, time periods, and data sources. Depending on the research question, the research team focused the analytic approach to align with program-specific parameters and available data. Each chapter contains a section that describes the limitations specific to the analysis and the mitigation strategies that were used to address the limitations.

Despite these limitations, the studies presented in this report provide a new platform from which to view the impact of CMS measures on specific patient populations and settings, efficiency of care processes, and quality health outcomes. Together, study findings support the development of innovative approaches for the future impact assessment of CMS quality measures.

Non-CMS entities sometimes use or promote the same measures as CMS, e.g., NCQA and private employers' use of HEDIS measures, The Leapfrog Group, Buying Value Initiative, and AHRQ patient safety measures. This has implications for the extent to which changes in processes of care or health outcomes can be solely attributed to CMS's deployment of measures.

Reference List

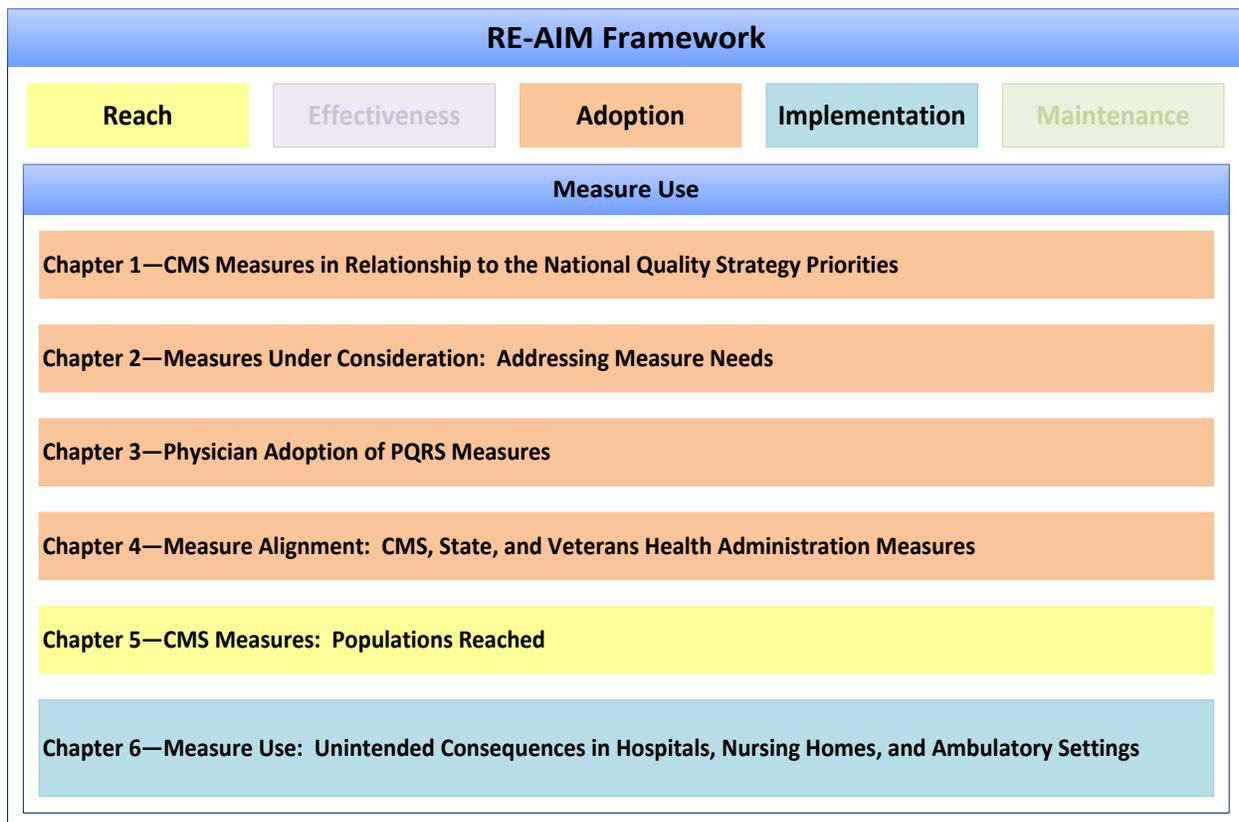
- (1) The Patient Protection and Affordable Care Act, Pub. L. No. 111–148, § 3014(b), 124 Stat. 119, 385–87 (2010) (as amended at 42 U.S.C. § 1395aaa-1).
- (2) Centers for Medicare & Medicaid Services. *National Impact Assessment of Medicare Quality Measures*. Baltimore, MD: Centers for Medicare & Medicaid Services. March 2012.
- (3) U.S. Department of Health and Human Services. *Report to Congress: National Strategy for Quality Improvement in Health Care*. Washington, DC: U.S. Dept of Health and Human Services; 2011
- (4) Centers for Medicare & Medicaid Services. *CMS Strategy: The Road Forward, 2013-2017*. Baltimore: MD: Centers for Medicare & Medicaid Services; March 2013.
- (5) Centers for Medicare & Medicaid Services. *CMS Quality Strategy - 2013 - Beyond*. Baltimore, MD: Centers for Medicare & Medicaid Services; November 18, 2013.
- (6) Dolin RH, Goodrich K, Kallem C, Alschuler L, Holtz P. Setting the standard. *J AHIMA*. 2014;85(1):42-48.
- (7) Chassin MR, Loeb JM, Schmaltz SP, Wachter RM. Accountability measures—using measurement to promote quality improvement. *N Engl J Med*. 2010;363(7):683-688.
- (8) Institute of Medicine of the National Academies. *Best Care at Lower Cost: The Path to Continuously Learning Health Care in America—Recommendations*. Washington, DC: Institute of Medicine of the National Academies; September 2012.
- (9) Centers for Medicare & Medicaid Services. CMS Measures Inventory. *Centers for Medicare & Medicaid Services*. May 21, 2014. Available at: <http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/Downloads/CMS-Measures-Inventory-.zip>. Accessed August 20, 2014.
- (10) Damberg C, Sorbero M, Lovejoy S et al. *An Evaluation of the Use of Performance Measures in Health Care*. Santa Monica, CA: National Quality Forum; 2011.
- (11) Conway PH, Mostashari F, Clancy C. The future of quality measurement for improvement and accountability. *JAMA*. 2013;309(21):2215-2216.
- (12) Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *Am J Public Health*. 1999;89(9):1322-1327.
- (13) National Research Council (U.S.), Committee on Health Impact Assessment, National Research Council (U.S.), Board on Environmental Studies and Toxicology. *Improving Health in the United States: The Role of Health Impact Assessment*. Washington, DC: National Academies Press; 2011.

- (14) Mindell JS, Boltong A, Forde I. A review of health impact assessment frameworks. *Public Health*. 2008;122(11):1177-1187.
- (15) Gaglio B, Shoup JA, Glasgow RE. The RE-AIM framework: a systematic review of use over time. *Am J Public Health*. 2013;103(6):e38-46.
- (16) Kessler R, Glasgow RE. A proposal to speed translation of healthcare research into practice: dramatic change is needed. *Am J Prev Med*. 2011;40(6):637-644.
- (17) Brown B. *Delphi Process: A Methodology Used for the Elicitation of Opinions of Experts*. Santa Monica, CA: The RAND Corporation; February 1968.

Part 1—Overview of CMS Quality Measure Use: Reach, Adoption, and Implementation

The first section of this report provides a high-level overview of measure use in the Centers for Medicare & Medicaid Services (CMS) quality measurement programs. The analyses of measure use presented in this section examine the extent to which CMS measures reflect national health priorities, as described in the National Quality Strategy (NQS) (Chapters 1 and 2); engage physicians, states, and the Veterans Health Administration in the use of CMS quality measures (Chapters 3 and 4); represent the Medicare population (Chapter 5); and have resulted in unintended consequences for patient care (Chapter 6). This broad perspective allows policymakers, researchers, and practitioners to reflect on the evolution and past progress of CMS measurement programs and to anticipate future directions for enhancing measure use to achieve the three aims of the NQS: Better Care, Healthy People/Healthy Communities, and Affordable Care. The topics examined and the aspects of the RE-AIM Framework addressed in each chapter of this section are presented in Figure P-1. The research question or questions are stated at the beginning of each chapter, and each chapter contains a table designating the programs and settings.

Figure P-1: Measure Use Chapter Titles Based on Research Questions



Chapter 1—CMS Measures in Relationship to the National Quality Strategy Priorities



Question on Adoption

What is the progress on addressing the National Quality Strategy (NQS) priorities as assessed by the Centers for Medicare & Medicaid Services (CMS) quality measures?

Abstract

Background: The U.S. Department of Health and Human Services (HHS) published the National Quality Strategy (NQS) in 2011 to provide a framework for coordinating quality measure development, implementation, and maintenance efforts. The NQS identified six priorities as key areas on which to focus resources and to achieve the aims of Better Care, Healthy People/Healthy Communities, and Affordable Care. The Centers for Medicare & Medicaid Services (CMS) Quality Strategy, published in November 2013, aligns with the NQS. Each of the NQS priorities has become a goal in the CMS Quality Strategy. To facilitate mapping of measures, CMS has established six measure domains associated with the NQS priorities. The goal of this research is to determine the distribution of CMS-implemented quality measures addressing the NQS priorities and associated domains and to illustrate change in the distribution of measures across each domain over time. This analysis builds on a prior study presented in the *National Impact Assessment of Medicare Quality Measures* report in 2012.

Methods: This descriptive analysis includes 822 unique current and finalized Medicare and Medicaid measures as of December 31, 2013, in 25 CMS quality measurement programs. A team of three clinically trained quality measure experts categorized the measures to one or more measure domains, according to the rules presented in the *HHS Decision Rules for Categorizing Measures of Health, Health Care Quality, and Health Care Affordability*. Using these decision rules, the three reviewers achieved 80 percent agreement on the classification of the measures to the domains. When the reviewers identified more than one measure domain, they designated a primary domain and listed the others as secondary. Cross-sectional frequency analyses at a single point in time were conducted to evaluate the extent to which the quality measures mapped to the domains and to calculate the number of

measures addressing multiple domains. A longitudinal analysis compared the changes in the number of measures in each domain between 2006 and 2013, during which HHS published its NQS in 2011 and CMS published its Quality Strategy in 2013.

Results: The number of measures in use in the CMS quality measurement programs has increased from 119 in 2006 to 822 in 2013, and CMS improved the balance of quality measures addressing the measure domains between 2006 and 2013. The increase in the number of measures is due in part to an increase in CMS quality measurement programs, which expanded from five to 25 between 2006 and 2013. Approximately 50 percent of the currently used or finalized CMS measures (n=822) apply to the Effective Treatment domain. The Care Coordination and Affordable Care priorities are the least represented with less than 7 percent of the measures for each of those domains. Approximately 40 percent of the measures address more than one priority, and approximately 2 percent of the measures address four domains.

Conclusions: Results from this study support the efforts by CMS to achieve an effective and efficient number of measures in each domain. To continue to address the objectives of the CMS Quality Strategy and national healthcare aims, CMS will need to maintain a parsimonious set of measures that improve healthcare quality, decrease the burden to providers, and fill measure gaps in underrepresented measure domains.

Background

The Patient Protection and Affordable Care Act (ACA) requires the Secretary of Health and Human Services to “establish a national strategy to improve the delivery of health care services, patient health outcomes, and population health.”¹ The National Quality Strategy (NQS) was first published in 2011 and was intended to provide a framework for coordinating the efforts of governmental, voluntary, and private sector quality improvement efforts.² In pursuit of this goal, the NQS sets out three broad aims:

- ◆ **Better Care:** Improve the overall quality by making healthcare patient-centered, reliable, accessible, and safe.
- ◆ **Healthy People/Healthy Communities:** Improve the health of the U.S. population by supporting proven interventions that address behavioral, social, and environmental determinants of health, in addition to delivering higher quality care.
- ◆ **Affordable Care:** Reduce the cost of quality healthcare for individuals, families, employers, and government.²

To advance these aims, the NQS focuses on six priorities.² The goals and objectives articulated in the Centers for Medicare & Medicaid Services (CMS) Quality Strategy, published in November 2013, reflect the six priorities of the NQS.³ Additionally, CMS has mapped the NQS priorities to six measure domains, which have been abbreviated throughout the report (Table 1-1).

Table 1-1: National Quality Strategy and CMS Quality Strategy

National Quality Strategy Priorities	CMS Quality Strategy Goals and Objectives	Measure Domains (Abbreviated)
1. Making care safer by reducing the harm caused in the delivery of care	Goal 1: Make care safer by reducing harm caused in the delivery of care <ul style="list-style-type: none"> ◆ Improve support for a culture of safety ◆ Reduce inappropriate and unnecessary care ◆ Prevent or minimize harm in all settings 	Patient Safety (Safety)
2. Ensuring that each person and family are engaged as partners in their care	Goal 2: Strengthen person and family engagement as partners in their care <ul style="list-style-type: none"> ◆ Ensure all care delivery incorporates patient and caregiver preferences ◆ Improve experience of care for patients, caregivers, and families ◆ Promote patient self-management 	Patient and Family Engagement (Patient Engagement)
3. Promoting effective communication and coordination of care	Goal 3: Promote effective communication and coordination of care <ul style="list-style-type: none"> ◆ Reduce admissions and readmissions ◆ Embed best practices to manage transitions to all practice settings ◆ Enable effective healthcare system navigation 	Care Coordination (Care Coordination)
4. Promoting the most effective prevention and treatment practices for the leading causes of mortality, starting with cardiovascular disease	Goal 4: Promote effective prevention and treatment of chronic disease <ul style="list-style-type: none"> ◆ Increase appropriate use of screening and prevention services ◆ Strengthen interventions to prevent heart attacks and strokes ◆ Improve quality of care for patients with multiple chronic conditions ◆ Improve behavioral health access and quality care ◆ Improve perinatal outcomes 	Clinical Process/ Effectiveness (Effective Treatment)
5. Working with communities to promote wide use of best practices to enable healthy living	Goal 5: Work with communities to promote best practices of healthy living <ul style="list-style-type: none"> ◆ Partner with and support federal, state, and local public health improvement efforts ◆ Improve access within communities to best practices of healthy living ◆ Promote evidence-based community interventions to prevent and treat chronic disease ◆ Increase use of community-based social services support 	Population/Public Health (Healthy Communities)
6. Making quality care affordable for individuals, families, employers, and governments by developing and spreading new healthcare delivery models (Affordable Care)	Goal 6: Make care affordable <ul style="list-style-type: none"> ◆ Develop and implement payment systems that reward value over volume ◆ Use cost analysis data to inform payment policies 	Efficient Use of Healthcare Resources (Affordable Care)

CMS identified four foundational principles that guide the Agency's actions toward the CMS Quality Strategy goals and objectives. The four principles are:

1. Eliminate racial and ethnic disparities.
2. Strengthen infrastructure and data systems.
3. Enable local innovations.
4. Foster learning organizations.

A strategic objective of CMS is to strengthen alignment of quality measures and the associated public reporting programs with the NQS. The aligned strategies are intended to improve patient outcomes and reduce the burden of measure reporting.⁴

The CMS Quality Strategy describes how CMS programs align with the priorities of the NQS.³ Examples include the following:

- ◆ CMS implemented the Hospital-Acquired Condition Reduction Program (HAC Reduction Program) to address the goal of making care safer by preventing or minimizing harm in the hospital setting.
- ◆ In support of the eligibility requirements for QIO contracts,^{viii,5} CMS is focusing on patient-centered care by supporting Quality Innovation Network-Quality Improvement Organization (QIN-QIO) initiatives, such as the Everyone with Diabetes Counts Program, which encourages patients and their families to take active roles in the care patients receive.
- ◆ The Hospital Readmission Reduction Program (HRRP) promotes effective communication and coordination of care by focusing on reducing hospital readmissions within 30 days after discharge.
- ◆ CMS promotes and supports many effective prevention and treatment programs for chronic diseases, including the Million Hearts Initiative, the Surviving Sepsis Campaign, Healthy People 2020, and other screening and treatment programs for leading causes of mortality and morbidity.
- ◆ CMS strives to make care affordable by implementing programs, such as the Hospital Value-Based Purchasing Program (Hospital VBP Program), which determines hospital payments based on performance on clinical processes of care, outcomes, and patient experiences.

CMS uses quality measurement as a lever to drive improvement on the national healthcare aims and priorities. With the proliferation of quality measures, it is critical to align measures across programs and across the public and private sectors to improve quality and reduce providers' burden. Therefore, it is essential to categorize measures according to domains aligned with the NQS priorities, which are similar to CMS goals. In 2013, in partnership with the National Quality Forum (NQF), HHS generated decision rules for healthcare agencies and measure developers to use when assigning new and existing quality measures to the six measure domains. These HHS Decision Rules (*HHS Decision Rules for Categorizing Measures of Health, Health Care Quality, and Health Care Affordability* in Appendix 1-1) helped standardize the application and interpretation of NQS priorities and assignment of measures to the domains. CMS staff and leadership participated in the creation and adoption of this set of decision rules, which CMS contractors use when categorizing measures by the measure domains. CMS strives to create a parsimonious set of measures that addresses the applicable domains and measure topics, using the fewest possible number of measures, which would minimize provider burden.

By applying the HHS Decision Rules to categorize the CMS measures by the measure domains, the research team examined the measure distribution by measure domain across and within CMS quality measure programs. The findings have policy and practice implications for strengthening CMS measures, so that quality measurement works as intended to advance progress on CMS and NQS priorities.

^{viii} The most recent final rule (78 FR 75198, Dec 10, 2013) added the following language that organizations must “demonstrate the ability to actively engage beneficiaries, families, and consumers, as applicable, in case reviews as set forth in §475.102; and/or quality improvement initiatives as set forth in §475.103” in order to be QIOs.

Objectives

Four objectives guided the analysis:

1. Determine the number of CMS measures that address each measure domain aligned with the NQS priorities.
2. Conduct a gap analysis to identify measure domains with relatively fewer measures.
3. Evaluate the extent to which the number of measures addressing each domain has changed over time between 2006 and 2013.
4. Examine the extent to which measures address multiple domains.

Methods

Measures and Data Sources

These descriptive analyses include 119 measures used in five CMS quality measurement programs in 2006 and 822 unique current and finalized measures used in 25 CMS quality measurement programs as of December 31, 2013. The research team used the CMS measures inventory to identify measures for inclusion. Appendix i-4 contains a hyperlink to a list of measures used in this analysis. This list includes the NQF endorsement status, NQF number if endorsed, and both the measure title used by the CMS program and the measure title used by NQF. Table 1-2 lists the included programs.

Table 1-2: Programs Included in Analysis

Setting	Program	Abbreviation
 Hospital	Hospital Inpatient Quality Reporting Program	Hospital IQR Program
	Hospital Value-Based Purchasing Program	Hospital VBP Program
	Hospital Readmissions Reduction Program	HRRP
	Hospital-Acquired Condition Reduction Program	HAC Reduction Program
	Medicare and Medicaid Electronic Health Record (EHR) Incentive Program for Eligible Hospitals and Critical Access Hospitals	EHR EH
	Hospital Outpatient Quality Reporting Program	Hospital OQR Program
	Ambulatory Surgical Center Quality Reporting Program	ASCQR Program
	Inpatient Psychiatric Facility Quality Reporting Program	IPFQR Program
	Prospective Payment System-Exempt Cancer Hospitals Quality Reporting Program	PCHQR Program
 Ambulatory	Physician Quality Reporting System	PQRS
	Medicare Electronic Prescribing Incentive Program	eRx Incentive Program
	Physician Feedback Program	Physician Feedback Program
	Medicare and Medicaid Electronic Health Record (EHR) Incentive Program for Eligible Professionals	EHR EP
	Medicare Shared Savings Program	MSSP
	Physician Compare	Physician Compare
	Medicare Part C (Display or Star Ratings when applicable)	Part C
	Medicare Part D (Display or Star Ratings when applicable)	Part D
	Core Set of Children's Health Care Quality Measures for Medicaid and CHIP (Child Core Set)	Medicaid Child
Core Set of Health Care Quality Measures for Adults Enrolled in Medicaid (Medicaid Adult Core Set)	Medicaid Adult	

Table 1-2: Programs Included in Analysis

Setting	Program	Abbreviation
 Post-Acute	Nursing Home Quality Initiative	NHQI
	Home Health Quality Reporting Program	HH QRP
	End-Stage Renal Disease Quality Incentive Program	ESRD QIP
	Hospice Quality Reporting Program	HQRP
	Inpatient Rehabilitation Facilities Quality Reporting Program	IRFQR Program
	Long-Term Care Hospitals Quality Reporting Program	LTCHQR Program

Analysis

The research team conducted a two-stage analysis to evaluate the number and distribution of measures in each of the measure domains across the CMS programs. First, a comprehensive list of measures (in use as of 2006 and in use or finalized for use as of December 31, 2013) was identified and assigned to each of the domains. Then, a series of descriptive analyses were conducted to address the four objectives.

Reviewers assigned measures to the domains using the HHS Decision Rules, developed by a cross-agency HHS workgroup in partnership with the NQF. Clinically trained quality measure experts with over 25 years of combined experience in clinical quality measurement independently reviewed each measure and assigned the measure to one or more of the domains. When the reviewers identified more than one domain, they designated a primary domain and listed the others as secondary. The three reviewers achieved an original agreement rate of 80 percent for the classification of measures to the domains. For each measure where the raters disagreed, the three reviewers discussed the classification until they achieved consensus on the classification. If they could not reach consensus, they consulted an additional clinical quality measure expert to reach a final resolution.

Objective 1—Determine the Number of CMS Measures That Address Each Measure Domain

The research team determined the frequency of measures addressing each domain (i.e., primary domain) by and across programs. Measures included were in use or finalized for use as of December 31, 2013. Each measure was counted once when calculating the number of CMS measures related to each measure domain, even if the measure was used in more than one program. In addition, within each specific program, the research team determined the number of measures relating to each domain.

Objective 2—Conduct a Gap Analysis to Identify Domains With Relatively Fewer Measures

The gap analysis determined how extensively the in-use and finalized measures address each of the domains and whether the domains are under- or overrepresented, based on a relative comparison to the distribution of measures empirically identified across other programs. The research team noted measure gaps if a domain was relatively less represented, compared with other domains in the current implementation of measures. The program purpose was considered when identifying gaps, since each domain may not be relevant to each of the programs. Measures in use or finalized as of December 31, 2013, were included.

Objective 3—Evaluate the Extent to Which the Number of Measures Addressing Each Domain Has Changed Over Time Between 2006 and 2013

The research team conducted a program-specific analysis of this objective for programs that started in 2006 or earlier by comparing the total number of measures reported in 2006 with the total number of measures reported in 2013 for each priority area. The results demonstrate the expansion in measure assignment to the measure domains within the quality measurement programs from 2006 to 2013.

Objective 4—Examine the Extent to Which Measures Address Multiple Domains

The analysis for the fourth objective evaluated the extent to which measures capture more than one measure domain. The research team counted the number of measures addressing multiple domains, and as in Objective 1, counted each unique measure once, though it may be included in more than one program. Measures in use or finalized for use as of December 31, 2013, were included and measures with a minimum of three years of data between 2006 and 2012 were included in a trend analysis by domain, presented in Chapter 7—CMS Measure Trends in Performance and Disparities.

Limitations

The research team encountered limitations in conducting the analyses. First, measures were assigned to domains using the HHS Decision Rules, which include an element of subjectivity. To minimize this, three independent, clinical reviewers applied the rules, discussed areas of disagreement, and reached a consensus. When the classifications in this study were compared with two external research teams, the percentage of agreement ranged between 64 percent and 77 percent, so the majority of measures were classified to the same domain across research teams, supporting internal consistency with the analysis. In addition to subjectivity, the agreement rates could have been affected by the differences in secondary classifications of the measures, since the HHS Decision Rules focus on assigning a primary domain and are not designed to provide guidance on secondary classifications.

Second, the HHS Decision Rules include a narrow definition of Affordable Care measures, which affected the number of measures assigned to this domain. Over-use measures are classified in the Safety domain if a knowledgeable patient would perceive the procedure as having significant risk, whereas previous analyses categorized over-use measures as Affordable Care.

Third, measures in the individual CMS quality measurement programs previously may have been categorized using different guidelines than the HHS Decision Rules. Those categories may not match the categories in this report. For example, the research team assigned the readmission measures to a primary domain of Care Coordination, according to the HHS Decision Rules. Other programs may have classified the readmission measures as Safety, Effective Treatment, or Affordable Care measures. The measure domains discussed in this report reflect the research team's assessment and not the original domains assigned by CMS when it adopted these measures in rulemaking.

Fourth, the purpose of a program may also affect the assignment of measures across the domains. For example, the HAC Reduction Program focuses on improving patient safety; thus, the program consists exclusively of Safety measures.

Fifth, CMS quality measurement programs that are relatively new have few measures in use. For example, the first reporting cycle for the Hospice Quality Reporting Program (HQRP), the Long-Term Care Hospitals Quality Reporting Program (LTCHQR Program), and the Inpatient Rehabilitation Facilities Quality Reporting Program (IRFQR Program) started in Fiscal Year 2014. These programs have 10⁶, 9⁷, and 5⁸ measures, respectively, finalized for implementation as of August 2014 for use in their respective programs. The number of measures addressing each domain in these new programs may be small compared with other programs.

Finally, measures used in multiple programs could result in artificially large numbers of measures in the measure domains. To mitigate this issue, the research team conducted the analysis on two levels. The first level, cross-program analyses, included unique measures. In the second level, program-specific analyses, the measures included in each program were counted. For example, *Central Line-Associated Blood Stream Infections (CLABSI)*^{ix} is included in the Hospital IQR Program, the Hospital VBP Program, and the HAC Reduction Program. For program-specific analyses, this measure counts as one measure in each program. For cross-program analyses, this measure counts as one unique measure.

The results are not intended to suggest that each of the domains should include the same percentage of measures. This analysis is descriptive and summarizes the number of measures addressing each domain. The analysis neither evaluates the effectiveness of the measures nor accounts for the relative contribution of each measure toward the attainment of CMS goals. However, CMS quality measurement programs can use the gaps identified in conjunction with an evaluation of the effectiveness of the current measures to inform decisions regarding future measures.

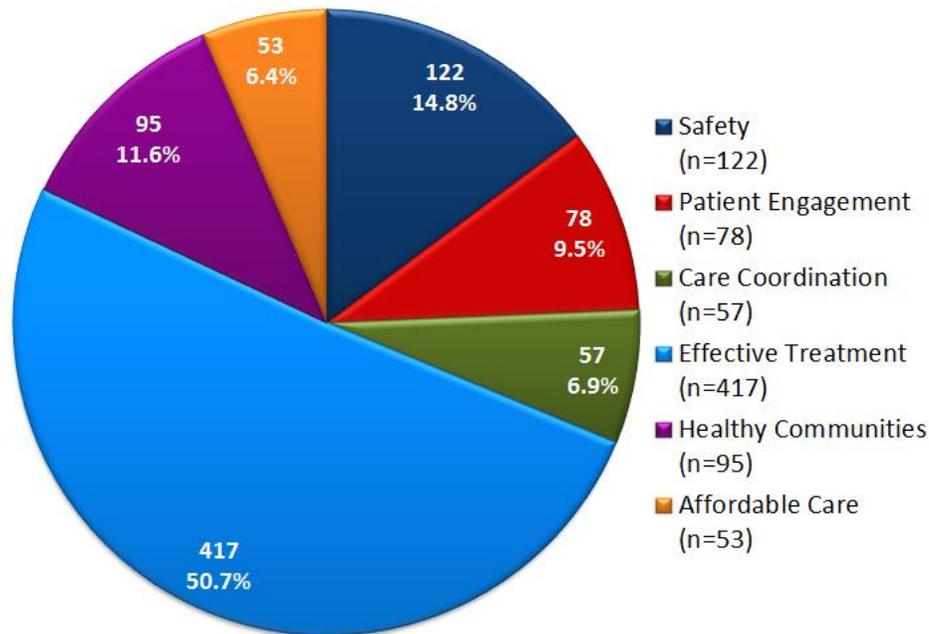
Results

Objective 1—Determine the Number of CMS Measures That Address Each Domain

As of December 31, 2013, 822 unique measures were in use or finalized for use in the 25 CMS quality measurement programs in this analysis. Two-thirds of the measures map to the Effective Treatment or Safety domains. Approximately 50 percent of those measures relate to the Effective Treatment domain; 14.8 percent of measures address Safety. The Care Coordination and Affordable Care domains are the least-represented domains, with 6.9 percent and 6.4 percent of the measures, respectively. Figure 1-1 shows the number and percent of measures relating to the domains across the CMS reporting programs.

^{ix} NQF-endorsed title: *National Healthcare Safety Network (NHSN) Central Line-Associated Bloodstream Infection (CLABSI) Outcome Measure* (NQF #0139).

Figure 1-1: Number of Unique Current Measures Addressing Each Measure Domain as of December 2013 (n=822)



The number of measures in each CMS program varies, as does the number of measures relating to each domain. Table 1-3 shows the distribution of measures addressing the measure domains for each CMS quality measurement program. Of the 822 unique measures, 317 measures (38.6 percent) are used in more than one program. Program-specific results (also shown in Table 1-3) include the total number of measures in each program, the sum of which is greater than the 822 unique measures shown in Figure 1-1, because the measures are used in multiple programs.

The ambulatory care CMS programs have the largest number of quality measures, with the Physician Feedback Program including 296 measures and the Physician Quality Reporting System (PQRS) including 284 measures. Due to the alignment of measures between the Physician Feedback Program and PQRS, 65.5 percent and 65.8 percent of the measures, respectively, map to the Effective Treatment domain.

The hospital-based programs also have a high percentage of measures addressing the Effective Treatment domain:

- ◆ 48.3 percent of EHR EH measures.
- ◆ 42.3 percent of Hospital Outpatient Quality Reporting Program (Hospital OQR Program) measures.
- ◆ 41.5 percent of Hospital IQR Program measures.
- ◆ 34.8 percent of Hospital VBP Program measures.

The hospital-based programs have a strong focus on Safety measures:

- ◆ 47.8 percent of Hospital VBP Program measures.

- ◆ 32.3 percent of Hospital IQR Program measures.
- ◆ 27.6 percent of EHR EH measures.
- ◆ 15.4 percent of Hospital OQR Program measures.

Of note, the Hospital-Acquired Condition Reduction Program (HAC Reduction Program) focuses on making care safer; therefore, each of the measures in this program addresses the Safety domain.

Table 1-3: Number and Percentage of Current Measures Addressing Each Domain by CMS Program (Measures May be Used in Multiple Programs)

CMS Program	Safety	Patient Engagement	Care Coordination	Effective Treatment	Healthy Communities	Affordable Care	Total
Hospital IQR Program	21 32.3%	5 7.7%	9 13.8%	27 41.5%	1 1.5%	2 3.1%	65 100.0%
Hospital VBP Program	11 47.8%	2 8.7%	0	8 34.8%	1 4.3%	1 4.3%	23 100.0%
HRRP	0	0	5 100.0%	0	0	0	5 100.0%
HAC Reduction Program	6 100.0%	0	0	0	0	0	6 100.0%
EHR EH	8 27.6%	3 10.3%	2 6.9%	14 48.3%	1 3.4%	1 3.4%	29 100.0%
Hospital OQR Program	4 15.4%	0	2 7.7%	11 42.3%	0	9 34.6%	26 100.0%
ASCQR Program	7 63.6%	0	0	2 18.2%	0	2 18.2%	11 100.0%
IPFQR Program	3 37.5%	0	3 37.5%	2 25.0%	0	0	8 100.0%
PCHQR Program	9 50.0%	1 5.6%	0	7 38.9%	0	1 5.6%	18 100.0%
PQRS	31 10.9%	17 6.0%	15 5.3%	187 65.8%	18 6.3%	16 5.6%	284 100.0%
eRx Incentive Program	0	0	1 100.0%	0	0	0	1 100.0%
Physician Feedback Program	32 10.8%	17 5.7%	16 5.4%	194 65.5%	18 6.1%	19 6.4%	296 100.0%
EHR EP	3 3.9%	4 5.2%	2 2.6%	48 62.3%	16 20.8%	4 5.2%	77 100.0%
MSSP	0	2 7.1%	3 10.7%	16 57.1%	7 25.0%	0	28 100.0%
Physician Compare	0	1 3.8%	2 7.7%	16 61.5%	7 26.9%	0	26 100.0%
Part C Display	1 6.3%	5 31.3%	1 6.3%	3 18.8%	2 12.5%	4 25.0%	16 100.0%
Part C Star Ratings	0	11 28.9%	3 7.9%	15 39.5%	6 15.8%	3 7.9%	38 100.0%
Part D Display	3 23.1%	3 23.1%	0	1 7.7%	0	6 46.2%	13 100.0%
Part D Star Ratings	1 5.6%	10 55.6%	0	2 11.1%	0	5 27.8%	18 100.0%

Table 1-3: Number and Percentage of Current Measures Addressing Each Domain by CMS Program (Measures May be Used in Multiple Programs)

CMS Program	Safety	Patient Engagement	Care Coordination	Effective Treatment	Healthy Communities	Affordable Care	Total
Medicaid Child	2 8.7%	1 4.3%	2 8.7%	2 8.7%	16 69.6%	0	23 100.0%
Medicaid Adult	2 7.7%	4 15.4%	3 11.5%	11 42.3%	6 23.1%	0	26 100.0%
NHQI	11 33.3%	2 6.1%	0	4 12.1%	16 48.5%	0	33 100.0%
HH QRP	9 10.5%	9 10.5%	9 10.5%	53 61.6%	6 7.0%	0	86 100.0%
ESRD QIP	2 12.5%	2 12.5%	0	12 75.0%	0	0	16 100.0%
HQRP	0	3 30.0%	0	7 70.0%	0	0	10 100.0%
IRFQR Program	3 60.0%	0	1 20.0%	0	1 20.0%	0	5 100.0%
LTCHQR Program	7 77.8%	0	1 11.1%	0	1 11.1%	0	9 100.0%

Objective 2—Conduct a Gap Analysis to Identify Domains with Relatively Fewer Measures

The Care Coordination and Affordable Care domains have the fewest number of unique measures, which are concentrated in the Hospital OQR Program, Medicare Part C Display, Medicare Part D Star Ratings, and Medicare Part D Display programs. As shown in Table 1-3, six of the 25 CMS quality measurement programs, i.e., Hospital IQR Program, EHR EH, PQRS, Physician Feedback Program, Medicare and Medicaid Electronic Health Record (EHR) Incentive Program for Eligible Professionals (EHR EP), and Part C Display, have measures in each domain. Three CMS quality measurement programs have measures that relate to one domain due to the specific focus of the program:

- ◆ The HRRP consists entirely of Care Coordination measures, since the program focuses on reducing readmissions.
- ◆ The HAC Reduction Program consists entirely of Safety measures, since the program focuses on reducing hospital-acquired conditions.
- ◆ The Medicare Electronic Prescribing Incentive Program (eRx Incentive Program) consists of one measure in the Care Coordination domain.

Certain CMS quality measurement programs do not include measures associated with each of the domains (Table 1-3):

- ◆ Five of the 25 CMS quality measurement programs do not have measures in the Effective Treatment domain.
- ◆ Six of the 25 CMS quality measurement programs do not have measures in the Safety domain.

- ◆ Eight of the 25 CMS quality measurement programs do not include measures that relate to the Patient Engagement domain.
- ◆ Nine of the 25 CMS quality measurement programs do not have measures in the Care Coordination domain.
- ◆ Eleven of the 25 CMS quality measurement programs do not have measures in the Healthy Communities domain.
- ◆ Fourteen of the 25 CMS quality measurement programs do not address the Affordable Care domain.

As part of the pre-rulemaking process, each year the NQF and the Measure Applications Partnership (MAP) review the quality measures under consideration for use in CMS quality measurement programs and provide input to HHS and CMS. The MAP considers gaps in the measure domains prior to making recommendations. Detailed information regarding the measures under consideration and the MAP process is presented in Chapter 2—Measures Under Consideration: Addressing Measure Needs.

Objective 3—Evaluate the Extent to Which the Number of Measures Addressing Each Domain Has Changed Over Time Between 2006 and 2013

The number of measures in use in CMS quality measurement programs in 2013 was more than six times the number of measures used in these programs in 2006 (Table 1-4).

Table 1-4: Number and Percentage of Unique Measures Addressing Each Measure Domain Between 2006 and 2013

Measure Domain	2006 (5 programs)		2011 (22 programs)		2013 (25 programs)		Direction of Change in Percentage From 2006 to 2013
	n	%	n	%	n	%	
Safety	14	11.8%	111	15.0%	122	14.8%	↑
Patient Engagement	5	4.2%	63	8.5%	78	9.5%	↑
Care Coordination	5	4.2%	37	5.0%	57	6.9%	↑
Effective Treatment	86	72.3%	420	56.9%	417	50.7%	↓
Healthy Communities	7	5.9%	71	9.6%	95	11.6%	↑
Affordable Care	2	1.7%	36	4.9%	53	6.4%	↑
Total	119		738		822		

The increase in the number of measures reflects an increase in the number of CMS quality measurement programs from five to 25 between 2006 and 2013. Also, as the CMS quality measurement programs have matured and diversified, the distribution of measures in the domains has changed. The proportion of measures addressing the Effective Treatment domain has decreased from 72.3 percent to 50.7 percent, whereas the percentage of Patient Engagement measures doubled between 2006 and 2013 and the percentage of Affordable Care measures quadrupled. Under the ACA, the pre-rulemaking process began in 2011 with measures under consideration by CMS for programs in 2012 and beyond. Measures in use in 2011 represent the baseline for the pre-rulemaking process. Chapter 2—Measures Under Consideration:

Addressing Measure Needs provides discussion on the pre-rulemaking process and the distribution of measures relative to the measure domains.

Table 1-5 shows the number of measures in each domain in 2006 and 2013 for the CMS quality measurement programs in existence in 2006, as reflected on the CMS Measures Inventory, which the research team selected as the primary source for comparison. According to that database, the Hospital IQR Program had 21 measures as of 2006 and 65 measures as of 2013.^{x,9,10,11} Of the 21 measures in 2006, 11 measures were removed by 2013, including eight measures relating to the Effective Treatment domain and three measures addressing the Patient Engagement domain. Each of the Nursing Home Quality Initiative (NHQI) measures used in 2006 was retired with the transition from Minimum Data Set (MDS) 2.0 to MDS 3.0; however, many of the same measure concepts were addressed in similar measures in MDS 3.0. Of the 54 Home Health Quality Reporting Program (HH QRP) measures available in 2006, 44 were included in 2013, and 10 had been retired (nine in the Effective Treatment domain and one in the Care Coordination domain). The 2006 End-Stage Renal Disease Quality Incentive Program (ESRD QIP) and Medicare Part C measures are included in the 2013 measures.

Table 1-5: Number of Measures Addressing Each Measure Domain for Five CMS Quality Measurement Programs in 2006 and 2013

CMS Program	Year	Safety	Patient Engagement	Care Coordination	Effective Treatment	Healthy Communities	Affordable Care	Total
Hospital IQR Program	2006	2	4	0	15	0	0	21
	2013	21	5	9	27	1	2	65
NHQI	2006	7	0	0	10	4	0	21
	2013	11	2	0	4	16	0	33
HH QRP	2006	5	0	4	45	0	0	54
	2013	9	9	9	53	6	0	86
ESRD QIP	2006	0	0	0	3	0	0	3
	2013	2	2	0	12	0	0	16
Part C ^{xi}	2006	0	1	1	13	3	2	20
	2013	1	16	4	18	8	7	54

Objective 4—Examine the Extent to Which Measures Address Multiple Domains

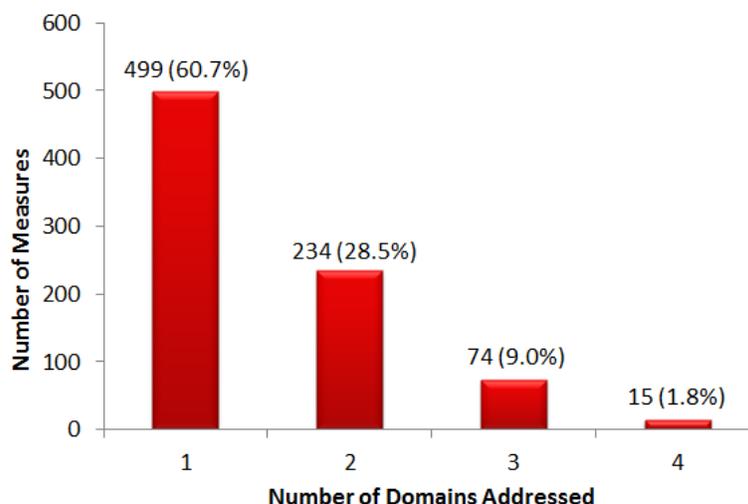
Measures that address more than one domain are of particular interest to CMS, since measures in multiple domains could be key drivers of healthcare quality. Measures addressing multiple domains help reduce provider burden and promote efficient use of resources, since fewer measures are

^x These 65 measures include: (1) 57 measures finalized for the FY 2016 payment determination and subsequent years in the 2013 IPPS Final Rule (Medicare Program; Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long Term Care Hospital Prospective Payment System and Fiscal Year 2014 Rates; Quality Reporting Requirements for Specific Providers; Hospital Conditions of Participation; Payment Policies Related to Patient Status; Final Rule. Fed Regist. 2013;78(160):50805); (2) 7 measures removed in that same rule for the FY 2016 payment determination and subsequent years (Fed Regist. 2013;78(160):50782); and (3) the 3-Item Care Transition Measure (CTM), NQF #0228—a measure typically included as part of the Patients' Experience of Care Measures/HCAHPS survey, but was treated by the research team as one measure for the purposes of this report. A suspended measure, Immunization for Pneumonia (IMM-1), was not included.

^{xi} Since Medicare Part C was not divided into the Part C Display and Part C Star Ratings programs in 2006, the Part C Display and Part C Star Ratings programs have been combined to compare 2006 to 2013.

required to achieve NQS goals. Measures corresponding to multiple domains may be adopted by other CMS programs, which could improve alignment across programs. As shown in Figure 1-2, 39.3 percent of the measures are in at least two domains (28.5 percent in two domains, 9.0 percent in three domains, and 1.8 percent of the measures in four domains). Of the 15 measures that are associated with four domains, eight are readmission measures. Readmission measures are efficient measures of quality, since they address aspects of healthcare related to patient safety, care coordination, effective treatment, and affordable care. Reviewers assigned the readmission measures to Care Coordination as the primary domain, according to the HHS Decision Rules.

Figure 1-2: Number of Unique Current Measures Addressing Multiple Domains as of December 2013



Identifying measures that map to the domains is necessary to understand the measure alignment to domains in the CMS quality measurement programs and thus gaps in domains. Beyond this analysis, examining the performance of the measures in each domain allows for an understanding of the effect of quality measures on NQS priorities. (These results are available in Chapter 7—CMS Measure Trends in Performance and Disparities.)

Discussion

CMS quality measurement programs have 822 unique measures currently in use or finalized for use. The physician-based programs have the largest number of measures. Given the wide variety of professionals (physicians, dentists, chiropractors, physician assistants, nurse practitioners, clinical psychologists, nutritionists, audiologists, and therapists) and medical specialties in the CMS clinician programs, these programs require a large number of measures, so providers can select appropriate measures for their practices.

Measures address each of the measure domains, as classified according to the HHS Decision Rules, although the measure assignment to domain varies across CMS quality measurement programs. Effective Treatment and Safety are the most prevalent domains with most CMS programs having a large percentage of measures in the Effective Treatment domain. Measures in the domains of Patient Engagement, Care Coordination, Healthy Communities, and Affordable Care are less prevalent.

Six programs, Hospital IQR, EHR EH, PQRS, Physician Feedback, EHR EP, and Part C Display, have measures in each domain. Other CMS programs have a narrow focus of quality measurement due to the statutory requirements; thus, the measures in these programs may relate to specific domains. For example, the HAC Reduction Program is limited to Safety measures, and the HRRP with a focus on reducing readmissions has measures corresponding to the Care Coordination domain. Programs that have been recently implemented, such as the LTCHQR Program or the IRFQR Program, may not have measures addressing each domain because the programs are new and started with a small number of measures. Measures addressing additional domains may be added as the programs mature and the number of measures increases.

Of the five programs that do not have measures in the Effective Treatment domain, the research team considers the HAC Reduction Program, HRRP, and the eRx Incentive Program to each focus on a specific domain (i.e., HAC focuses on Patient Safety, HRRP focuses on Care Coordination, and eRx focuses on Care Coordination). The other two programs, the LTCHQR Program and the IPFQR Program, are new programs with small numbers of measures.

Six CMS quality measurement programs do not have measures in the Safety domain. HRRP and the eRx Incentive Program each focus on a specific goal that does not include Safety measures, and HQRP is new with a small number of measures. The lack of a measure in the Safety domain for the Part C Star Ratings program indicates this could be a priority area for new measure development. The Medicare Shared Savings Program (MSSP) and the Physician Compare program are relatively new programs, which could account for not having measures in the Safety domain.

Of the eight CMS quality measurement programs that do not have measures in the Patient Engagement domain, the HAC Reduction Program, HRRP, and the eRx Incentive Program have a specific focus, as previously described, that does not include measures addressing the Patient Engagement domain. The IPFQR Program, IRFQR Program, and LTCHQR Program are new programs with small numbers of measures. Neither the Hospital OQR Program nor the Ambulatory Surgical Center Quality Reporting Program (ASCQR Program) has measures that address patient engagement, which represents a gap in measurement.

Part D Star Ratings or Display measures do not address the Care Coordination domain. While the Hospital VBP Program and the HAC Reduction Program do not have measures in the Care Coordination domain, other hospital-based programs have measures addressing Care Coordination, including HRRP, which specifically addresses reducing readmissions and consists exclusively of Care Coordination measures.

Eleven of the programs do not have measures in the Healthy Communities domain, and 14 of the 25 CMS quality measurement programs do not have measures in the Affordable Care domain. None of the programs in the post-acute care settings has measures that address Affordable Care, indicating this is an important area for measure development.

Evaluating the change in the distribution of measures across the domains over time shows that CMS has improved the proportion and balance of quality measures addressing the domains between 2006 and 2013. The Effective Treatment domain, with a large portfolio of measures decreased from 72 percent in 2006 to 51 percent by 2013. The percentage of measures

addressing the Patient Engagement and Healthy Communities domains increased the most between 2006 and 2013.

As the number of measures increases, the burden on providers to collect data on the measures may increase. The identification of high-impact measures that relate to multiple domains may help to reduce provider burden and to achieve effective outcomes. Using data that CMS is collecting for other purposes can further reduce burden. In addition to reducing provider burden and facilitating efficient use of resources, measures that correspond to multiple domains could be high impact because they address several focus areas of healthcare quality.

Results should be interpreted with caution because the HHS Decision Rules clearly define the guidelines for classifying measures to the primary domain, but not for classifying secondary domains. Classification of secondary domains may be more subjective than the classification of the primary domain. This analysis focused solely on how CMS measures address the six measure domains that align with the NQS priorities and the corresponding CMS Quality Strategy goals. Further analysis is necessary to determine how measures address the four foundational principles of the CMS Quality Strategy, including healthcare disparities.

Conclusions

Key Findings

Results from this study support the efforts by CMS to achieve an effective and efficient number of measures in each domain and the importance of continuing this effort. About 50 percent of the measures currently in use or finalized for use in CMS quality measurement programs address the Effective Treatment domain with most CMS programs having a measure in this domain. Nearly 15 percent of the measures relate to the Safety domain. The Care Coordination and Affordable Care domains are the least-represented domains, with over one-half of the CMS quality measurement programs lacking Affordable Care measures. Across the CMS quality measurement programs, nearly 10 percent of the measures are associated with patient engagement; however, as CMS increases the focus on patient-centered care, the number of measures addressing patient engagement may increase.

CMS improved the balance of quality measures addressing the NQS priorities between 2006 and 2013, despite differences in the absolute number of measures in each category and the variation of measures in each measure domain across programs. The number of measures implemented in the individual CMS quality measurement programs ranges from one measure in the eRx Incentive Program to nearly 300^{xii} measures in the PQRS. The number of measures currently in use or finalized for use has increased from 119 in 2006 to 822 in 2013, a more than six-fold increase, which is due in part to the expansion in the number CMS quality measurement programs from five to 25 between 2006 and 2013.

Nearly 40 percent of the measures currently in use or finalized for use in the CMS quality measurement programs address more than one domain, with approximately 2 percent of the

^{xii} This number includes individual measures and measures belonging to measure groups in the PQRS.

measures in four domains. The measures that correspond to multiple domains have the potential to be important indicators of healthcare quality because they address several different aspects of the quality of care delivered.

Actions to Consider

CMS will continue analyzing the distribution of measures related to the goals of the CMS Quality Strategy and national healthcare aims to identify less frequently represented domains. These domains should be considered during new measure development for CMS reporting programs.

Action to Consider: Focus on Affordable Care and Care Coordination domains as high priorities for new measure development. Novel approaches to measure design and development may be required to address these critical gaps effectively. These approaches include hybrid data sources, e.g., claims and electronic health record data, and shared accountability between providers.

Action to Consider: Implement Patient Engagement measures in the institutional outpatient settings (Hospital OQR Program and ASCQR Program). CMS is already working on the development and implementation of Consumer Assessment of Healthcare Providers and Systems (CAHPS®)^{xiii} surveys for the ASC and Emergency Department settings. Prioritize new measures addressing Patient Engagement in the newer CMS quality measurement programs (LTCHQR Program, IRFQR Program, and IPFQR Program).

Action to Consider: Ensure the applicable measure domains are addressed as new programs, such as LTCHQR Program, IRFQR Program, IPFQR Program, HQRP, and Physician Compare, implement new measures.

Analyses should further investigate the role of measures addressing multiple domains to determine if these measures are high-impact measures that could be tools to increase healthcare quality, while minimizing provider burden. As CMS strives to create a parsimonious set of measures for each of the applicable domains, measures that address multiple domains may help achieve these goals, while minimizing provider burden.

Action to Consider: Evaluate measures addressing multiple domains and examine their potential roles in reducing provider burden, promoting efficient use of resources, and increasing harmonization across programs. If high-impact measures are identified, implement these measures in the applicable CMS quality measurement programs.

Action to Consider: Review programs that have a large percentage of measures in specific domains, e.g., Effective Treatment, to confirm that each measure addresses an important aspect of healthcare quality, is directly linked to patient outcomes, has variation in performance, and minimizes the burden on providers.

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

Reference List

- (1) The Patient Protection and Affordable Care Act, Pub. L. No. 111–148, § 3011, 124 Stat. 119, 378–80 (2010).
- (2) U.S. Department of Health and Human Services. *Report to Congress: National Strategy for Quality Improvement in Health Care*. Washington, DC: U.S. Dept of Health and Human Services. 2011.
- (3) Centers for Medicare & Medicaid Services. *CMS Quality Strategy: 2013–Beyond*. Baltimore, MD: Centers for Medicare & Medicaid Services. November 18, 2013.
- (4) Centers for Medicare & Medicaid Services. *CMS Strategy: The Road Forward, 2013-2017*. Baltimore, MD: Centers for Medicare & Medicaid Services. March 2013.
- (5) Eligibility Requirements for QIO Contracts. 42 CFR, Part 475:504 (2009).
- (6) Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long Term Care Hospital Prospective Payment System and Fiscal Year 2015 Rates; Final Rule. *Fed Regist.* 2014;79(163):50485-50496.
- (7) Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long Term Care Hospital Prospective Payment System and Fiscal Year 2015 Rates; Final Rule. *Fed Regist.* 2014;79(163):50286-50305.
- (8) Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long Term Care Hospital Prospective Payment System and Fiscal Year 2015 Rates; Final Rule. *Fed Regist.* 2014;79(163):45908-45914.
- (9) Revision to Hospital Inpatient Prospective Payment Systems-2007 FY Occupational Mix Adjustment to Wage Index; Implementation; Final Rule. *Fed Regist.* 2006;71(160):48033-48034.
- (10) Medicare Program; Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long-Term Care Hospital Prospective Payment System and Fiscal Year 2013 Rates; Hospitals' Resident Caps for Graduate Medical Education Payment Purposes; Quality Reporting Requirements for Specific Providers and for Ambulatory Surgical Centers. *Fed Regist.* 2012;77(170):53530-53533.
- (11) Medicare Program; Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long Term Care; Hospital Prospective Payment System and Fiscal Year 2014 Rates; Quality Reporting Requirements for Specific Providers; Hospital Conditions of Participation; Payment Policies Related to Patient Status; Final Rule. *Fed Regist.* 2013; 78 (160): 50780.

Chapter 2—Measures Under Consideration: Addressing Measure Needs



Question on Adoption

How are the gaps in measures being addressed through the pre-rulemaking process?

Abstract

Background: The Patient Protection and Affordable Care Act requires the U.S. Department of Health and Human Services (HHS) to publish a list of quality and efficiency measures that are being considered for use in Medicare programs (the pre-rulemaking process). HHS has contracted with the National Quality Forum to convene the Measure Applications Partnership (MAP), a multi-stakeholder group, to provide input to HHS and the Centers for Medicare & Medicaid Services (CMS) on a list of quality and efficiency measures being considered for adoption through rulemaking in the Medicare program. This study examines how the list of measures considered for CMS quality programs addresses the National Quality Strategy (NQS) priorities and the outcomes of the pre-rulemaking process.

Methods: In this descriptive study, the research team compiled the lists of measures during the pre-rulemaking process from 2011 to 2013. Three reviewers assigned an NQS priority for each measure, according to the rules presented in the January 6, 2014 document, *HHS Decision Rules for Categorizing Measures of Health, Health Care Quality, and Health Care Affordability*. Using these decision rules, the reviewers achieved 80 percent agreement on the classification of the measures according to the priorities. To understand the impact of the pre-rulemaking process on the portfolio of CMS measures, the research team compared the distribution of measures across the measure domains. The research team also compiled recommendations for adoption that each measure received from the MAP and examined federal documents to identify the implementation status of measures in CMS programs. The research team used qualitative techniques to analyze text from federal documents to generate key themes about the reasons CMS had for adopting measures not recommended by the MAP. Frequency analyses showed how measures submitted to the MAP addressed NQS priorities over the three-year period.

Results: CMS submitted 874 unique measures to the MAP between 2011 and 2013. The number of measures in domains relating to each of the six NQS priorities varied each year with a consistent positive trend in the Affordable Care domain from 2011 to 2013. The number of measures submitted for each Medicare program also varied each year, with ranges of two to 153 in 2011, five to 281 in 2012, and one to 158 in 2013. Between 2011 and 2013, CMS adopted 208 measures that were under consideration for implementation in Medicare programs. The number of measures implemented in CMS programs for Care Coordination, Affordable Care, and Healthy Communities increased. On average, the MAP supported (i.e., supported, supported direction, and conditionally supported) 52 percent (n=673) and did not support 35 percent (n=458) of the measures submitted by CMS during the three pre-rulemaking cycles. The MAP did not provide individual measure recommendations for 13 percent (n=162) of the measures submitted during the same time period. Out of 446 measures that received MAP support in 2011 and 2012, 180 measures (40 percent) were published in final rules as of December 2013 for implementation in CMS programs. Between 2011 and 2012, 282 measures received a “Do Not Support” recommendation from the MAP, of which 28 (10 percent) were published in final rules for implementation in CMS programs. CMS diverged from MAP recommendations and included measures not recommended by the MAP for implementation in Medicare programs because these 28 measures addressed topics important to CMS, evaluated aspects of quality for which there are a limited number of measures, and allowed specialty providers to participate in Medicare quality measurement programs.

Conclusions: CMS has worked to align the list of measures it brings to the MAP for consideration to fill gaps in NQS priority areas. CMS has implemented a small proportion of measures not recommended by the MAP to achieve the goals of the NQS and allow for broader provider participation.

Background

The public posting and solicitation of input from stakeholders on the measures under consideration for the Centers for Medicare & Medicaid Services (CMS) programs prior to the formal national rulemaking process is relatively new. Section 3014 of the Patient Protection and Affordable Care Act (ACA) created a new section 1890A of the Social Security Act (the Act), which requires the establishment of a federal “pre-rulemaking process” for the selection of quality and efficiency measures for use in specific Medicare programs. These categories of measures are described in section 1890(b)(7)(B) of the Act.¹

To comply with the ACA Section 3014 pre-rulemaking process, CMS publishes annually by December 1 of each year a list of quality and efficiency measures it is considering for adoption through rulemaking in the Medicare program. The ACA also requires a consensus-based entity to convene multi-stakeholder groups to provide input to the U.S. Department of Health and Human Services (HHS) on the list of measures put forth by CMS for use in the Medicare

program. The National Quality Forum (NQF) is currently the consensus-based entity under contract by HHS to fulfill this requirement.

The Measure Applications Partnership (MAP), the multi-stakeholder group convened by the NQF, as required under section 1890A of the ACA, provides input to HHS on the pre-rulemaking measures. The Secretary is required to submit a list of measures for MAP input during each pre-rulemaking cycle for the CMS Medicare programs identified as meeting the criteria described in the ACA. Accordingly, the NQF-convened MAP reviews the quality and efficiency measures CMS is proposing in rulemaking. CMS considers the NQF-convened MAP recommendations when proposing measures for its programs.

The MAP uses a set of criteria to make recommendations regarding the selection of measures for the CMS programs (Appendix 2-1).² The measure selection criteria focus on ensuring that the high-quality measures selected for CMS program implementation address the National Quality Strategy (NQS) aims,^{2,3} fill critical measurement gaps, and increase alignment. These criteria help the MAP identify characteristics associated with ideal measure sets used for public reporting and payment programs. The measure selection criteria provide general guidance but do not supersede program-specific statutory and regulatory requirements.

Prior to the ACA, HHS and CMS received input on measures for their programs by issuing a Notice of Proposed Rule Making (NPRM) for one healthcare program at a time. This process allows stakeholders and other interested parties to provide comments before the issuance of a final rule specifying the measures selected for program implementation.

Because rules are program-specific, a comprehensive assessment of measures in use by the federal government is not easily accomplished. The new pre-rulemaking process, however, provides a fully coordinated vision for performance measurement across CMS. While the notice and public comment rulemaking continues, the pre-rulemaking process enables a systematic review of measures that CMS is considering for all applicable programs simultaneously.

The *National Impact Assessment of Medicare Quality Measures* (2012 Impact Report) examined the list of measures submitted to the MAP in 2011, the first pre-rulemaking cycle.⁴ Due to limitations in the data, the report focused primarily on the number of measures and the NQS priority domains addressed by the measures. Additionally, NQF analyzed the measure list to examine the uptake of MAP recommendations by HHS.⁵⁻⁷ For example, the NQF concordance analysis of MAP recommendations with measures chosen by CMS for implementation, according to the final rules published in 2012, showed that concordance varied among the programs. The findings indicated that concordance between MAP recommendations and CMS implementation plans was 70 percent or greater for the majority of the programs reviewed by the MAP.⁸ In 2013, the NQF analysis also showed over 90 percent concordance between the MAP “Do Not Support” recommendation and the decision of CMS not to finalize measures.⁹

Evidence from previous analyses did not provide information regarding how the pre-rulemaking process has improved alignment between the measures used in CMS programs and the NQS priorities. This study expands on these prior analyses by determining the extent to which the measures submitted to the MAP addressed NQS priorities in 2011, 2012, and 2013. This study

also includes an examination of measures not supported by the MAP but finalized for implementation in CMS programs.

Objectives

Three objectives guided the analysis to examine how gaps in measures are being addressed through the pre-rulemaking process. The objectives of this study are:

1. Determine the number of measures under consideration that addressed each measure domain for each year (2011, 2012, and 2013) and analyze the trends in addressing the gap areas.
2. Summarize MAP recommendations and characterize the reasons for implementing measures that the MAP did not support.
3. Determine the extent to which the pre-rulemaking process improved alignment between measures implemented by CMS and the six measure domains.

Methods

Measures and Data Sources

This is a descriptive study of the measures submitted by CMS to the MAP for input in December 2011, 2012, and 2013 and publicly posted on the NQF website. The data included measures for the 17 CMS programs described in Table 2-1. A hyperlink to a list of all measures used in this study is provided in the Appendix i-4. The list includes the NQF endorsement status, NQF number if endorsed, and both the measure title used by the CMS program and the measure title used by NQF.

Although measures are submitted to the MAP for 17 programs described in Table 2-1, programs may elect not to submit measures for the pre-rulemaking process in a given year, depending on program needs for additional measures and the availability of new measures addressing those needs. In addition, CMS may include measures for programs not mandated by the ACA in the list of measures submitted to the MAP. For example, measures proposed for the Medicare and Medicaid Electronic Health Record (EHR) Incentive Program for Eligible Professionals (EHR EP) have been historically included because of the alignment of the EHR EP measures with other programs, such as the Physician Quality Reporting System (PQRS) and the Hospital Inpatient Quality Reporting Program (Hospital IQR Program). CMS is also required to resubmit measures to the MAP, if the measures have undergone substantial changes in their specifications and/or are proposed for use in a different CMS program.

Table 2-1 contains the list of CMS programs for which measures are submitted to the MAP.

Table 2-1: Programs Included in Analysis

Setting	Program	Abbreviation
 Hospital	Hospital Inpatient Quality Reporting Program	Hospital IQR Program
	Hospital Value-Based Purchasing Program	Hospital VBP Program
	Hospital Readmissions Reduction Program	HRRP
	Hospital-Acquired Condition Reduction Program	HAC Reduction Program
	Hospital Outpatient Quality Reporting Program	Hospital OQR Program
	Ambulatory Surgical Center Quality Reporting Program	ASCQR Program
	Inpatient Psychiatric Facility Quality Reporting Program	IPFQR Program
	Prospective Payment System-Exempt Cancer Hospital Quality Reporting Program	PCHQR Program
 Ambulatory	Physician Quality Reporting System	PQRS
	Physician Feedback Program	Physician Feedback Program
	Medicare Shared Savings Program	MSSP
	Physician Compare	Physician Compare
 Post-Acute	Home Health Quality Reporting Program	HH QRP
	End-Stage Renal Disease Quality Incentive Program	ESRD QIP
	Hospice Quality Reporting Program	HQRP
	Inpatient Rehabilitation Facilities Quality Reporting Program	IRFQR Program
	Long-Term Care Hospitals Quality Reporting Program	LTCHQR Program

Analysis

Objective 1—Determine the Number of Measures Under Consideration That Addressed Each Measure Domain for Each Year (2011, 2012, and 2013) and Analyze the Trends in Addressing the Gap Areas

The measures on each of the three measure lists were assigned an NQS priority, using the same methodology described in Chapter 1—CMS Measures in Relationship to the National Quality Strategy Priorities. A team of three clinically trained quality measure experts independently assigned each measure to one or more of the measure domains (Appendix 2-2). The reviewers assigned the measures to the domains according to the rules presented in the January 6, 2014 document, *HHS Decision Rules for Categorizing Measures of Health, Health Care Quality, and Health Care Affordability* (HHS Decision Rules), which is a set of consensus decision rules, established by a cross-agency HHS workgroup in partnership with NQF (Appendix 1-1). A senior clinical reviewer compared each reviewer's results to determine the amount of inter-rater agreement. Using the HHS Decision Rules, the three reviewers achieved an original agreement rate of 80 percent for the classification of measures to domains. The research team resolved disagreement among the raters by discussing the classification and reaching consensus. If they

did not achieve consensus, a senior clinical reviewer determined the final resolution. Upon completion, the research team submitted the final list of measures and their assigned domains to CMS for review.

CMS and the MAP used the six domains to categorize the measures for each of the three years. However, there was no uniform definition and methodology for assigning measures to NQS priorities prior to the development of the HHS Decision Rules. As a result, the categorizations used in this study may vary from published pre-rulemaking measure lists, MAP reports, and CMS documents. Although a measure may address more than one priority, this study used the priority that was the “best fit” for the measure.

Objective 2—Summarize MAP Recommendations and Characterize the Reasons for Implementing Measures That the MAP Did Not Support

The research team gathered MAP recommendations for each measure from the MAP pre-rulemaking reports published annually. The MAP evaluated each measure submitted for consideration and provided one of the following recommendations:

- ◆ Support—indicates the MAP recommended a measure for a CMS program in the current rulemaking cycle.
- ◆ Support Direction/Conditional Support—indicates the MAP recommended the measure be phased into a CMS program to allow time to address specific issues regarding the measure.
- ◆ Do Not Support—indicates the MAP did not recommend the measure for implementation in a CMS program. Measures with missing recommendations were included and noted in the analysis.

The research team reviewed *Federal Register* notices for calendar years 2012 and 2013 to determine the implementation status of select measures in CMS quality, reporting, and payment programs. Text from the *Federal Register* notices was analyzed to understand why CMS implemented measures that did not receive support from the MAP. The research team used thematic analysis to generate key themes found in *Federal Register* notices pertaining to implementation of measures submitted to the MAP. *Federal Register* final rules published from calendar year January 1, 2012, to December 31, 2013, were analyzed. A senior research team member closely examined federal documents and noted each instance where a measure under consideration not supported by the MAP was adopted in a CMS program. To discover common themes, the senior research team member read and categorized findings in the text by key phrases, words, or synonyms that CMS used in its rationale for using measures not supported by the MAP.

Objective 3—Determine the Extent to Which the Pre-Rulemaking Process Improved Alignment Between Measures Implemented by CMS and the Six Measure Domains

To understand the impact of the first two cycles of the pre-rulemaking process (2011 and 2012) on the portfolio of CMS measures, the research team compared the distribution of measures across the domains. The research team reviewed the final rules published in the *Federal Register* in 2012 and 2013 for the quality measurement programs included in the pre-rulemaking process (n=17) after the first two pre-rulemaking cycles. The measures in the final rules were compared

to the list of measures submitted to the MAP in 2011 and 2012. This analysis excludes the measures submitted to the MAP during 2013 because the implementation plans for these measures were included in final rules published after this study was concluded.

Using CMS measures in use as of 2011 as a baseline, the change in the number and proportion of measures implemented or finalized for programs as of December 31, 2013, was calculated. The research team assigned an NQS priority to a measure using the methodology described in Objective 1.

Limitations

Information regarding provider performance on the implemented measures that CMS submitted during the pre-rulemaking process will not be available for a few years. Therefore, the research team was not able to include the evaluation of the impact of the measures selected for implementation as an objective for this 2015 Impact Report. As data become available, future studies can evaluate the impact of these measures.

The analyses for Objectives 1 and 2 include the measures submitted to the MAP in 2011, 2012, and 2013. The analyses for this chapter were completed before the final rules were published in 2014; therefore, the analysis of the implementation of the MAP recommendations for the measures submitted in 2013 (Objective 3) could not be included in this report.

CMS may resubmit a measure in future years to the MAP due to substantial changes to the measure methodology. Substantial changes to a measure can result in a re-review by the MAP, as well as possibly the NQF Steering Committee for consensus review and evaluation, if the measure steward is seeking endorsement or if the measure was previously endorsed. A measure that previously received a “Do Not Support” recommendation may receive a “Support” recommendation when submitted in the succeeding years. For example, if a measure was previously not recommended because the measure was not consistent with current clinical practice guidelines, a revised version of the measure that is consistent with current clinical guidelines may receive support.

Measure domain classifications may not be consistent with other reports and analyses. To maintain consistency in assigning the domains, the research team used the HHS Decision Rules, which CMS and other HHS agencies currently use. In addition, while it may be appropriate to classify a measure under more than one domain, for the purposes of this report, the primary domain is used.

Other factors affect the number of measures included in programs that address NQS priorities. The optimal number of measures that should represent each priority is not known. Program policy is another factor that affects the number of measures implemented in programs. For example, one of the purposes of the Medicare and Medicaid Electronic Health Record (EHR) Incentive Program for Eligible Professionals is to encourage eligible professionals to use electronic health records for quality measurement through the use of incentive payments and downward payment adjustments. Therefore, the program includes a variety of measures for reporting by both general and specialty providers. If measures for a specialty provider, such as a pathologist, are lacking, CMS may opt to include pathology-related measures for operationalizing the program, even if these measures do not address a particular NQS priority gap.

The data used in the analysis had limitations because the MAP did not provide recommendations for certain measures. For example, in the second pre-rulemaking cycle, CMS published a list of 507 measures under consideration for the applicable programs on December 1, 2012. The MAP indicated in the recommendation report that it did not have sufficient information to provide a recommendation for a measure that was submitted for both the Ambulatory Surgical Center Quality Reporting Program (ASCQR Program) and Hospital Outpatient Quality Reporting Program (Hospital OQR Program).⁷ Additionally, the MAP did not provide recommendations for 27 measures submitted for the Physician Feedback Program (Physician Feedback) and Physician Compare. Rather than providing recommendations for these measures, the MAP developed guiding principles for applying these measures in clinician performance programs, which provided a foundation for selection of measures for these programs. Subsequently, the MAP conducted meetings after the publication of the pre-rulemaking report, where the MAP provided input for these measures. The research team reviewed MAP meeting documents and found MAP recommendations for 20 out of 27 measures submitted for Physician Feedback. MAP recommendations for nine measures submitted during 2012 were not found.

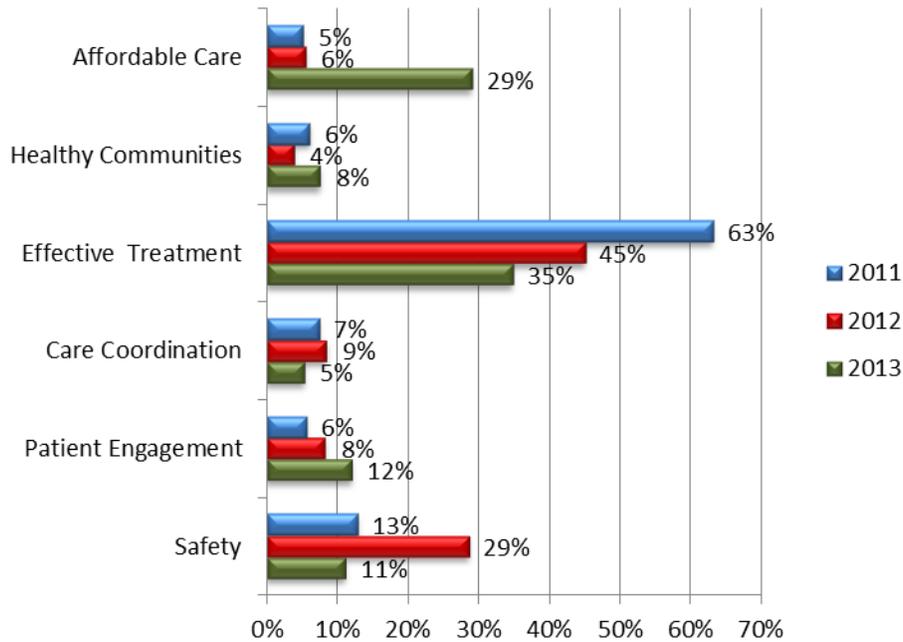
CMS changes and discussions with NQF about the 2013 measure list led to a number of measures (n=153) not reviewed for recommendations by the MAP. Therefore, the MAP did not provide recommendations concerning these measures that were originally submitted December 1, 2013.

Results

Objective 1—Determine the Number of Measures Under Consideration That Addressed Each Measure Domain for Each Year (2011, 2012, and 2013) and Analyze the Trends in Addressing the Gap Areas

CMS submitted to the MAP identical or aligned measures that are applicable for implementation in more than one program; 874 unique measures were submitted between 2011 and 2013. Of these measures, 217 measures were submitted in 2011, 433 measures were submitted in 2012, and 224 measures were submitted in 2013.

Figure 2-1 illustrates the percentage of measures by domain submitted to the MAP for consideration between 2011 and 2013. The proportion of measures that focus on Effective Treatment accounted for the largest share of the measures submitted in 2011, 2012, and 2013, although a clear, downward trend is noted between 2011 and 2013. Conversely, a positive trend is notable in the percentage of measures submitted for the Affordable Care (5 percent to 29 percent of measures submitted) and Patient Engagement (6 percent to 12 percent) domains from 2011 to 2013.

Figure 2-1: Percentage of Measures submitted to the MAP by Measure Domain (2011–2013)

Objective 2—Summarize MAP Recommendations and Characterize the Reasons for Implementing Measures That the MAP Did Not Support

MAP Recommendations for Measures Under Consideration

In the succeeding analyses, the research team identified the total number of measures by program (i.e., the same measure may be counted in more than one program) because the MAP recommendation for identical measures submitted in different programs may not be the same. In 2011, the MAP reviewed 233 measures submitted across 10 of 17 programs, with a range of five to 153 measures per program. In 2012, the MAP reviewed 504 measures submitted across 15 of 17 programs, with a range of two to 281 measures per program. Finally, in 2013, the MAP reviewed 556 measures submitted across 16 of 17 programs, with a range of one to 158 measures per program. From 2011 to 2013, the MAP reviewed a total of 1,293 measures, which includes duplicated counts of measures submitted for multiple CMS programs.

Table 2-2 shows a summary of MAP recommendations for measures that were reviewed in 2011, 2012, and 2013. The MAP supported (i.e., supported the direction of or conditionally supported) 108 measures (46 percent) in 2011, 338 measures (67 percent) in 2012, and 227 measures (41 percent) in 2013. Alternately, the MAP did not support a total of 125 measures (54 percent) in 2011, 157 measures (31 percent) in 2012, and 176 measures (32 percent) in 2013. In the report *Input on Measures Under Consideration by HHS for 2012 Rulemaking*, the MAP indicated that nearly 70 percent of the total number of measures that were not supported did not have enough measure information.¹⁰ Recommendations for nine measures (2 percent) submitted by CMS in 2012 and 153 measures (27 percent) in 2013 were not specifically provided by the MAP. (See

Appendices 2-3 through 2-5 for program-specific summaries of MAP recommendations for measures submitted for consideration by year.)

Table 2-2: Summary of MAP Recommendations by Year

Year	Support/Support Direction n (%)	Not Supported n (%)	Recommendation Not Provided n (%)	Number of Measures Submitted n (%)
2011	108 (46%)	125 (54%)	0	233 (100%)
2012	338 (67%)	157 (31%)	9 (2%)	504 (100%)
2013	227 (41%)	176 (32%)	153 (27%)	556 (100%)
Total	673 (52%)	458 (35%)	162 (13%)	1,293 (100%)

Table 2-3 presents a summary of the rationale provided by the MAP for not recommending measures for implementation. From 2011 to 2013, there was an increasing trend in the frequency for the MAP rationale “Better measure exists” and “Need to submit for endorsement.” The number of measures that the MAP did not recommend because better measures exist increased from eight measures in 2011 to 29 measures in 2013. For these measures, the MAP determined that a measure addressing a similar topic that better addresses the current needs of the program was available. The number of non-recommended measures that required submission for NQF endorsement consideration increased from one measure in 2011 to 34 measures in 2013.

The number of measures that had the rationale, “Does not address needs of the program,” was higher in 2012 (n=48) and 2013 (n=39) compared with 2011 (n=7). Across the three years, these measures comprised over 20 percent of the total number of measures that received a “Do Not Support” recommendation.

Table 2-3: Summary of MAP Rationale for “Do Not Support” Measures

MAP Rationale	2011	2012	2013	All Years
Specifications not provided	102	0	0	102
Does not address needs of program ^{xiv}	7	48	39	94
Rationale not given	1	74	2	76
Measure needs further development	0	0	66	66
Better measure exists	8	11	29	48
Need to submit for endorsement	1	8	34	43
Not endorsed	4	15	6	25
Reliability concerns	2	1	0	3
Total	125	157	176	458

^{xiv} The MAP evaluates the ability of an individual measure to meet the program needs, using guiding principles that complement the MAP Measure Selection Criteria and program-specific statutory and regulatory requirements. For example, the guidance for inclusion of measures in PQRS indicates to include outcome measures that are not already addressed by outcome measures in the program or are clinically relevant to specialties/subspecialties that do not currently have clinically relevant measures.

CMS Implementation of Measures With MAP Support^{xv}

The research team conducted a quantitative analysis of CMS rulemaking. This included a comparison of measures that received MAP support for program implementation and final rules published in the *Federal Register* in calendar years 2012 and 2013, which identified the concordance between the MAP recommendations and CMS implementation decisions for measures submitted during the pre-rulemaking process. From 2011 to 2012, CMS implemented 180 measures (40 percent) out of 446 measures that received MAP support for program implementation. Table 2-4 shows a summary of CMS implementation plans for measures supported by the MAP.

Table 2-4: Summary of CMS Measure Implementation Plans for Measures Supported by the MAP by Year

Year	Used in CMS Program n (%)	Not Used in CMS Program n (%)	Total
2011	80 (74%)	28 (26%)	108 (100%)
2012	100 (30%)	238 (70%)	338 (100%)
Total	180 (40%)	266 (60%)	446 (100%)

CMS Implementation of Measures Without MAP Support

Analyses of the MAP recommendations for the list of measures submitted for consideration from 2011 to 2012 were compared with the measures in final rules published in the *Federal Register* in calendar years 2012 and 2013 to identify discordance between the decision of the MAP not to support measures and the decision of CMS to implement non-supported measures. CMS finalized for implementation 28 measures that received a “Do Not Support” recommendation from the MAP. These measures accounted for 10 percent of the total number of measures (n=282) that received a “Do Not Support” recommendation from the MAP. Ninety percent of the measures that received a “Do Not Support” recommendation were not selected for implementation in CMS programs. Table 2-5 shows a summary of CMS implementation plans for measures not supported by the MAP.

Table 2-5: Summary of CMS Measure Implementation Plans for Measures Not Supported by the MAP by Year

Year	Used in CMS Program n (%)	Not Used in CMS Program n (%)	Total
2011	17 (13%)	108 (87%)	125 (100%)
2012	11 (7%)	146 (93%)	157 (100%)
Total	28 (10%)	254 (90%)	282 (100%)

For the 28 measures not supported by the MAP in 2011 and 2012, program-level analysis was conducted. Results for measures not supported by the MAP but finalized for implementation by CMS include Hospital VBP Program (one measure), PQRS (25 measures), and Hospital IQR Program (two measures). Appendix 2-3 and Appendix 2-4 present program-specific summaries of CMS implementation plans for 2011 and 2012 measures that did not receive MAP support.

^{xv} The discussion does not necessarily reflect the policies and rationales described by CMS when it adopted the measures in rulemaking.

CMS Rationale for Implementing Measures Not Supported by the MAP

The research team reviewed the *Federal Register* for final rules published for the measurement programs included in the analysis to determine the CMS rationale for implementing measures that the MAP did not recommend. Subsequently, the findings were categorized into themes including importance of measure topic, unavailability of alternative measures, and the need to include measures specific to medical specialty groups (e.g., surgery), which are summarized in Table 2-6:

Table 2-6: Rationale for Measures Implemented by CMS that were Not Supported by the MAP

Program	Importance of Measure Topic	Unavailability of Alternative Measures	Need to include Measures for Medical Specialty Groups
Hospital IQR	N/A	2	N/A
Hospital VBP	1	N/A	N/A
PQRS	16	1	9

In addition, the research team identified key activities that CMS conducted to overcome issues identified by the MAP, including the following:

Importance of Measure Topic

One of the themes that emerged from the review of CMS documents was importance of measure topic, which was identified through statements, such as “filling gaps” and “high-priority area,” that CMS used as rationale for finalizing measures not supported by the MAP. For example, 16 measures that were not recommended by the MAP due to unavailability of measure specifications were proposed for implementation in PQRS for 2013 because the measures filled key gaps.¹¹ Similarly, *Screening Colonoscopy Adenoma Detection Rate* was finalized for implementation in PQRS for 2014 because the measure addresses a “broad patient population for screening and detection of colorectal cancer and is medically significant in the measurement of utilizing preventive healthcare services.”¹²

Importance of measure topic was also a theme found during the review of CMS documents pertaining to the NQF-endorsed measure *PSI 90: Complication/Patient Safety for Selected Indicators Composite*^{xvi}, which CMS implemented as part of the Hospital VBP Program FY 2015 measure set. Although the MAP supported the use of the measure for the Hospital IQR Program, the MAP expressed concerns over the reliability of the measure, which uses Medicare data exclusively, and recommended that its use should not be linked to payment.¹⁰ However, CMS believed that this measure addresses an important topic—patient safety—and is appropriate for use in the Hospital VBP Program. To address concerns regarding reliability, CMS adopted a longer performance period to collect additional data for performance scoring. This resulted in the postponement of the implementation of NQF #0531 until Fiscal Year 2016 in the Hospital VBP Program.¹³

^{xvi} NQF-endorsed title: *Patient Safety for Selected Indicators* (NQF #0531).

Unavailability of Alternative Measures

Another theme that emerged from the review of CMS documents was unavailability of alternative measures. For example, CMS indicated that the measure *Shared Decision-Making: Trial of Conservative (Non-surgical) Therapy* was chosen to be implemented in PQRS because the number of measures that are focused on this key topic were limited.^{14, 11}

Similarly, CMS finalized two measures for the Hospital IQR Program that the MAP did not support¹⁵: *Mortality-30-Stroke: Hospital 30-day, All-Cause, Risk-Standardized Mortality Rate (RSMR) following a Stroke Hospitalization* and *READM-30-Stroke: Hospital 30-day, All-Cause, Risk-Standardized Readmission Rate (RSRR) following an Acute Ischemic Stroke Hospitalization*. Although these measures were not NQF-endorsed or MAP-supported, CMS considered other available measures that were endorsed and found no other feasible and practical measures on these topics.¹⁶ In the *MAP 2014 Recommendations on Measures for More Than 20 Federal Programs*, the MAP discussed the appropriateness of the CMS decision to implement these measures in the Hospital IQR Program and subsequently supported the stroke mortality and readmission measures for the Hospital IQR Program.¹⁵

Need to Include Measures for Medical Specialty Groups

The final theme that emerged from the review of CMS documents was the need to include measures for medical specialty groups. This theme was apparent in the nine measures finalized for PQRS for 2014 (Appendix 2-6). CMS included these measures because the measures allowed additional reporting opportunities for eligible professionals who previously had a limited number of measures available for reporting.

For example, the MAP did not support a group of measures applicable to bariatric procedures that CMS submitted for PQRS; however, CMS elected to implement these measures (PQRS measure numbers 354, 355, 356, 357, and 358)^{11, 17} as part of a general surgery measures group, which is applicable to several procedures including bariatric surgery.^{xvii} This decision supported the goal of CMS to provide ample reporting opportunities to eligible professionals, especially those who are unable to report other broadly applicable measures. Further, the addition of a general surgery measures group created reporting options (other than claims or registry) for the surgeons in this group.^{11, 17}

Objective 3—Determine the Extent to Which the Pre-Rulemaking Process Improved Alignment Between Measures Implemented by CMS and the Six Measure Domains

The research team reviewed the *Federal Register* for the quality measurement programs, which are included in the pre-rulemaking process (n=17) and determined that CMS adopted for implementation 208 measures that the MAP reviewed during the first two pre-rulemaking cycles. Table 2-7 shows the total number of measures submitted to the MAP in 2011 and 2012 that CMS

^{xvii} The procedures included in the general surgery group include Ventral Hernia, Appendectomy, AV Fistula, Cholecystectomy, Thyroidectomy, Mastectomy +/- Lymphadenectomy or Sentinel Lymph Node Biopsy (SLNB), Partial Mastectomy or Breast Biopsy/Lumpectomy +/- Lymphadenectomy or SLNB, Bariatric Laparoscopic or Open Roux-en Y Gastric Bypass, Bariatric Sleeve Gastrectomy, and Colectomy.

adopted for program implementation as of December 31, 2013, stratified by domain. The highest number of measures adopted by CMS addressed Effective Treatment (n=84), followed by Safety (n=50), Care Coordination (n=27), Affordable Care (n=21), Healthy Communities (n=17), and Patient Engagement (n=9).

Table 2-7: Number and Percentage of Measures Reviewed by the MAP and Adopted by CMS Stratified by Measure Domain (2011–2012)

Measure Domain	2011		2012		Total	
	n	%	n	%	n	%
Effective Treatment	46	47%	38	34%	84	40%
Safety	22	23%	28	25%	50	24%
Care Coordination	8	8%	19	17%	27	13%
Affordable Care	10	10%	11	10%	21	10%
Healthy Communities	7	7%	10	9%	17	8%
Patient Engagement	4	4%	5	5%	9	4%
Total	97	100%	111	100%	208	100%

Using 2011 as a baseline and 2013 as a comparator, Table 2-8 also illustrates the percent change in measures reviewed by the MAP and adopted by CMS from Baseline 2011 to 2013 for each domain. Comparatively the percentages of measures addressing Care Coordination, Healthy Communities, Affordable Care, and Patient Engagement increased, while the percentage of measures addressing Effective Treatment and Safety decreased from baseline to 2013. Proportionally, measures that address Effective Treatment accounted for over half of the measures in 2011 and 2013. Measures that address Care Coordination and Affordable Care accounted for the smallest percentage (6.7 percent and 6.9 percent, respectively) of the total number of measures in 2013.

Table 2-8: Summary of Number and Percentage of Measures by Measure Domain at Baseline 2011 and 2013

Measure Domain	Baseline 2011		2013		% Change
	n	%	n	%	
Care Coordination	31	4.5%	51	6.7%	2.2%
Healthy Communities	52	7.6%	71	9.3%	1.7%
Affordable Care	36	5.2%	53	6.9%	1.7%
Patient Engagement	58	8.4%	71	9.3%	1.5%
Safety	107	15.6%	115	15.0%	-0.6%
Effective Treatment	403	58.7%	405	52.9%	-5.8%
Total	687	100%	766	100%	

Discussion

This descriptive study examined the measures that CMS submitted to the MAP for input as part of the ACA 3014 pre-rulemaking process and the extent to which the measures addressed the measure domains, as reassigned by the research team for the purposes of this study. The pre-rulemaking process gives stakeholders and CMS the opportunity to examine the entire portfolio

of measures used across programs and allows for a comprehensive assessment of how CMS is responding to the priorities set forth in the NQS.

A high degree of concordance between MAP recommendations and CMS implementation for the measures submitted during the pre-rulemaking process was identified. Analysis of MAP decisions and final rules as of December 2013 indicated that CMS had not selected for implementation a majority (90 percent) of the measures that the MAP did not recommend during the pre-rulemaking process. CMS finalized 10 percent of the non-recommended measures for implementation in CMS programs. These results are consistent with the concordance analysis of MAP recommendations conducted by NQF.⁹ However, analysis of the CMS rationale for implementing measures not supported by the MAP revealed compelling reasons for adopting these measures in federal programs, including the importance of measure topic, unavailability of alternative measures, and the need to include measures for specialty groups. Additionally, 40 percent of the measures that the MAP recommended for program implementation have been finalized in CMS programs as of December 2013. The analysis in this 2015 Impact Report included measures finalized in rules after the first two pre-rulemaking cycles (2011 and 2012) were completed. To assess the impact of the pre-rulemaking process on CMS programs, this analysis can continue in future years as additional pre-rulemaking and rulemaking cycles are completed.

Findings from the analysis of the MAP rationale of non-recommended measures and CMS rationale for implementing non-recommended measures highlight the complex tasks involved in selecting potential measures for program implementation that meet the current goals of the program and the goals for aligning measures across CMS programs. Examination of the MAP rationale for not recommending measures showed an increasing trend in the number of measures receiving negative recommendations due to the existence of better measures and the need for NQF endorsement. Over one-third of the measures received a “Do Not Support” recommendation in 2013 from the MAP because the MAP determined the need for further measure development. This suggests that the MAP prefers evaluating measures that are more fully developed.

The MAP recommendations depend on the information the MAP receives during the pre-rulemaking process. To ensure the MAP has sufficient information to guide decisions, CMS consistently expanded the information provided to the MAP. CMS also streamlined operational aspects of the pre-rulemaking process in 2013 using Lean principles, which resulted in the elimination of processes that do not add value to the pre-rulemaking procedure and engaged federal and non-federal stakeholders much earlier in the process.

The MAP has made the following recommendations to aid in its evaluation of the measures submitted by CMS:

- ◆ Seek and utilize additional quantitative and qualitative information on measures.
- ◆ Ensure that both potential positive and negative impacts are evaluated.
- ◆ Consider a stronger focus on measures that address upstream health determinants of large populations.
- ◆ Look beyond general impact to variations in impact for different populations that may signal disparities.

- ◆ For selected measures, develop explicit hypotheses and/or estimates on the range of impact that can be evaluated against outcomes at a later time.¹⁸

In addition, the MAP is proposing that predictive modeling/forecasting methods be used to predict the future performance of the measures.¹⁸ Such a model would need to be developed and tested. Implementation of this recommendation is not likely to occur for several years. Until then, measures submitted to the MAP should include sound rationale, implementation and performance history, barriers to implementation, and evaluation of measure performance conducted by the developer or other organizations.

Conclusions

Key Findings

A positive trend was notable in the percentage of measures submitted to the MAP for consideration in both the Affordable Care and Patient Engagement domains from 2011 to 2013. This trend suggests that CMS was submitting measures to the MAP in domains that were proportionally lower than other domains within the portfolio. However, Care Coordination, which is also underrepresented, exhibited a negative trend in the percentage of measures.

The MAP recommended slightly more than half (51 percent) of the measures that were submitted between 2011 and 2013. Conversely, the MAP did not support 35 percent of the measures, and 14 percent of the measures did not receive a recommendation. The most frequent rationale for the MAP not supporting quality measures in the most recent rulemaking cycle (2013) included measure needs further refinement, measure does not address needs of the program, and the need to submit the measure for endorsement. These findings suggest the MAP prefers measures that have been fully specified and tested.

CMS implemented 40 percent of the measures that were supported by the MAP. CMS also implemented 28 measures not supported by the MAP, which represented approximately 10 percent of the total number of measures not supported by the MAP. In each instance, rationale was provided for implementation of the measures and the themes identified included importance of measure topic, unavailability of alternative measures, and the need to include measures specific to medical specialty groups (e.g., surgery) to provide opportunities for participation in quality measurement programs.

The portfolio of measures representing programs, which undergo the pre-rulemaking process, was more balanced proportionally in 2013 than in 2011. Programs with proportionally fewer measures increased, i.e., Care Coordination, Healthy Communities, Affordable Care, and Patient Engagement, and programs representing a larger share of the portfolio decreased, i.e., Patient Safety and Effective Care. This finding should be considered preliminary, as the amount of data is limited, and there may be a lag time between a recommendation from the MAP and the implementation of the measure into a rule.

Actions to Consider

CMS has made progress by proposing measures to the MAP to address underrepresented measure domains; however, additional quality measures are needed to address underrepresented areas, such as Care Coordination and Affordable Care.

Action to Consider: Submit measures to the MAP addressing underrepresented domains, such as Affordable Care and Care Coordination. Since these domains, as identified in Chapter 1, are underrepresented, opportunities exist for filling additional gaps with measures that address these domains.

CMS provides a concise and relevant summary of measure information and specifications to support the MAP decision-making process. The study findings suggest the MAP prefers more fully developed and tested measures. As CMS considers continued enhancement of the information provided to the MAP, a key component of this information should include evidence supporting the linkage between improved evidence-based processes and outcomes and the projected impact on health outcomes based on quality improvement scenarios.

Action to Consider: Include key findings from the measure business case analysis, when available, in the information provided to the MAP. A key component of the business case analysis is establishing the outcome/process linkage and modeling the quality improvement goals and objectives that CMS hopes to achieve by implementing the measure. This information would assist the MAP in understanding the potential impact of measures submitted.

Action to Consider: Consider the inclusion of measure developers in the MAP process to address measure-specific questions. Measure developers can support CMS staff, as needed, to address any measure-specific questions efficiently and further articulate the potential impact of measure implementation on outcomes. Measure developer input can be particularly valuable to ensure MAP members understand key technical details related to measure specifications and the key findings from measure testing.

This study was limited to analysis of the number of measures by measure domain and indicated that the CMS measure portfolio was becoming more proportionally balanced over time. However, while the relative proportion of measures is an important trend to monitor, future analysis could address the effectiveness of current measures.

Action to Consider: Evaluate both the proportional trends in the number of measures attributed to each domain, as well as classify and quantify the performance gap for measures currently in-use, e.g., determining measures that are “topped out.” Stratification of the number of measures by domain and performance gap may further assist CMS in targeting new measure development for underrepresented domains and domains where a large proportion of measures have limited room for improvement.

Reference List

- (1) The Patient Protection and Affordable Care Act, Pub. L. No. 111–148, § 3014(b), 124 Stat. 119, 385–87 (2010) (codified at 42 U.S.C. § 1395aaa-1).
- (2) National Quality Forum. MAP Measure Selection Criteria. MAP Task Forces. http://www.qualityforum.org/MAP_Task_Forces.aspx. Published October 15, 2013. Accessed September 24, 2014.
- (3) U.S. Department of Health and Human Services. *Report to Congress: National Strategy for Quality Improvement in Health Care*. Washington, DC: U.S. Dept of Health and Human Services. 2011.
- (4) Centers for Medicare & Medicaid Services. *National Impact Assessment of Medicare Quality Measures*. Baltimore, MD: Centers for Medicare & Medicaid Services; March 1, 2012.
- (5) Measure Applications Partnership. *Measure Applications Partnership Clinician Workgroup Web Meeting*. [PowerPoint]. Washington, DC: National Quality Forum; December 2, 2013.
- (6) Measure Applications Partnership. *Measure Applications Partnership Post-Acute Care/Long-Term Care Workgroup Web Meeting*. [PowerPoint]. Washington, DC: National Quality Forum; November 4, 2013.
- (7) National Quality Forum. *MAP Pre-Rulemaking Report: 2013 Recommendations on Measures Under Consideration by HHS*. Washington, DC: National Quality Forum; February 1, 2013.
- (8) National Quality Forum. *Uptake of MAP Recommendations in HHS Final Rules During 2012*. [PowerPoint]. Washington DC: National Quality Forum; 2013.
- (9) National Quality Forum. *CMS Uptake of MAP 2013 Pre-Rulemaking Recommendations*. [PowerPoint]. Washington DC: National Quality Forum; 2013.
- (10) National Quality Forum. *Input on Measures Under Consideration by HHS for 2012 Rulemaking*. Washington, DC: National Quality Forum; February 1, 2012.
- (11) Medicare Program; Revisions to Payment Policies under the Physician Fee Schedule, Clinical Laboratory Fee Schedule & Other Revisions to Part B for CY 2014. *Fed Regist.* 2013;78(139):43388-43420.
- (12) Medicare Program; Revisions to Payment Policies Under the Physician Fee Schedule, Clinical Laboratory Fee Schedule & Other Revisions to Part B for CY 2014. *Fed Regist.* 2013;78(237):74652-74653.
- (13) Medicare Program, Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long-Term Care Hospital Prospective Payment System and Fiscal Year

- 2013 Rates; Hospitals' Resident Caps for Graduate Medical Education Payment Purposes; Quality Reporting Requirements for Specific Providers and for Ambulatory Surgical Centers (Final Rule), 42 CFR Parts 412, 413, 424, and 476, Department of Health and Human Services, Centers for Medicare & Medicaid Services (2012).
- (14) Medicare Program; Revisions to Payment Policies Under the Physician Fee Schedule, Clinical Laboratory Fee Schedule and Other Revisions to Part B for CY 2014; Final Rule, 42 CFR Parts 405, 410, 411, 414, 423, and 425, Department of Health and Human Services, Centers for Medicare and Medicaid Services (2013).
- (15) National Quality Forum. *MAP 2014 Recommendations on Measures for More Than 20 Federal Programs*. Washington, DC; January 31, 2014.
- (16) Medicare Program, Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long-Term Care Hospital Prospective Payment System and Fiscal Year 2014 Rates; Quality Reporting Requirements for Specific Providers; Hospital Conditions of Participation; Payment Policies Related to Patient Status (Final Rule), 42 CFR Parts 412, 413, 414, 419, 424, 482, 485, and 489, Department of Health and Human Services, Centers for Medicare & Medicaid Services (2013).
- (17) Medicare Program; Revisions to Payment Policies Under the Physician Fee Schedule, Clinical Laboratory Fee Schedule & Other Revisions to Part B for CY 2014. *Fed Regist.* 2013. 78(237): 74647-74653.
- (18) National Quality Forum. *Measures Application Project*. Coordinating Committee In-Person Meeting. Washington, DC. October 3, 2012.

Chapter 3—Physician Adoption of PQRs Measures



Questions on Adoption

1. Which physician and patient characteristics are associated with physician participation in the voluntary Physician Quality Reporting System (PQRS)?
2. Which physician and patient characteristics are associated with the types of measures PQRS participants chose to report?

Abstract

Background: The Physician Quality Reporting System (PQRS) is a quality measure reporting program that has not had universal participation by eligible professionals (including physicians) who provide services under Medicare Fee-For-Service. The research team sought to identify characteristics that are associated with PQRS measure participation among eligible professionals (i.e., physicians and non-physician professionals); due to data limitations, non-physician eligible professionals were excluded from the analysis. The research team also sought to identify the key characteristics (such as specialty) that account for selection of particular measures to report among physicians participating in PQRS.

Methods: The research team used multiple logistic regression to show how PQRS participants and non-participants differed by patient population, practice, and physician characteristics, as well as to estimate the relationship between measure selection decisions by PQRS participants and multiple factors, including the physician's specialty and the proportion of the participant's patient populations eligible for particular measures. The study population consisted of 650,423 physicians who did not reassign their benefits to a hospital or group practice. These physicians submitted individual Medicare Part B physician claims in 2012 and were also included in Medicare Physician Compare in May 2013. The research team linked physician PQRS participation using 2012 clinician-level data on measure reporting. Patient population characteristics and total Medicare Part B payments to physicians (who did not reassign their benefits to a hospital or group practice) were identified using Medicare Master Beneficiary Summary File, Medicare Hierarchical Condition

Category scores, and Medicare Part B physician claims. Physician primary specialty was extracted via Physician Compare data.

Results: The baseline participation rate in PQRS among physicians in the sample was 41.3 percent in 2012. Compared with physicians in practices with 25 or more physicians, those in solo practice were less likely to participate in PQRS (-26.1 percentage points, p value <0.0001). Physicians with sicker populations were less likely to participate in PQRS; participation rates were 4.2 percentage points lower among physicians whose patients' average Hierarchical Condition Category (HCC) score was in the 90th percentile compared with those whose patients' average HCC score was in the 10th percentile. Participation rates were 5.9 percentage points lower among physicians with a non-white proportion in the 90th percentile compared with those with a non-white proportion in the 10th percentile. An increase in the annual Medicare Part B payments by \$10,000 over the average payment to physicians who did not reassign their benefits to a hospital or group practice increased the likelihood of physician participation by 0.5 percentage points. Not surprisingly, physicians were more likely to choose measures within their specialties than measures outside of their specialties. For example, compared with primary care physicians, cardiologists were much more likely to report measures related to coronary artery disease (+20.5 percentage points), while anesthesiologists (+70.2 percentage points) and general surgeons (+29.5 percentage points) were much more likely to select measures related to perioperative care.

Conclusions: Physician participation in PQRS appears to be strongly correlated with practice size. Among PQRS participants, self-reported physician specialty is also a strong predictor for measure selection decisions; most participating physicians reported measures appropriate for their specialties. Participants also chose measures that reflected their patient population, though measure selection was more strongly associated with physician specialty than the proportion of patients eligible for the measures. Physicians are participating in PQRS at higher rates and are choosing measures related to their patients' clinical conditions; in other settings, similar participation increases have translated into better care for Medicare beneficiaries. Further research is needed to determine the significance of physicians with sicker patients and physicians with higher proportions of non-white patients being less likely to participate and what factors might be precluding participation among physicians with these patient population characteristics.

Background

The Centers for Medicare & Medicaid Services (CMS) began implementing the Physician Quality Reporting System (PQRS) in 2007 to improve the quality of ambulatory care delivered to Medicare beneficiaries.¹ PQRS measures the performance of eligible professionals,^{xviii} including physicians, practitioners, and therapists.^{2,3} Ending in 2014, PQRS has provided eligible professionals bonuses of up to 2 percent of their total annual Part B Medicare Physician Fee Schedule payments, if they satisfactorily reported PQRS measures in a given year. As mandated by legislation authorizing PQRS, the financial incentives related to physician participation in PQRS changed in 2013; non-participation in 2013 will result in penalties levied in 2015.

Between 2007 and 2012, CMS increased the number of measures from which physicians could choose to report for PQRS from 74 to 266, thereby increasing the number of specialties able to report. The number of professionals eligible and able to participate in PQRS increased from less than 700,000 in 2006 to more than one million in 2012.² In contrast to nearly universal participation rates by hospitals in the CMS Hospital Inpatient Quality Reporting Program,⁴ a smaller proportion of eligible professionals have chosen to participate in PQRS. PQRS participation rates have increased 21 percent between 2007 (15 percent participation) and 2012 (36 percent participation), although the changing number of professionals eligible to report to PQRS makes it difficult to compare participation rates over time.² (Of note, due to data limitations, this analysis was limited to physicians [i.e., clinicians with MD or DO degrees].)

A limited number of studies have examined the association between participation in PQRS and the characteristics of participating clinicians. Federman and Keyhani analyzed differences between participants and non-participants in a sample of 4,934 physicians and found no differences among them, based on age, gender, specialty, or region.⁵ CMS reported that PQRS participation rates varied by region and that primary care clinicians participated at lower rates than specialists, such as anesthesiologists, radiologists, and emergency medicine physicians.^{2,6} Less variability in the PQRS participation rates across income levels and practice size was observed.² The research team hypothesized that practice size should have a substantial effect after accounting for other factors, such as specialty, because there might be economies of scale in quality reporting. For example, larger practices may have greater financial resources to hire additional staff to collect and report data for PQRS measures.⁷ The research team also hypothesized that characteristics of physicians' patient populations might influence their decisions to participate in PQRS, but no studies were identified that examined this question empirically. Finally, the research team hypothesized that physicians' choices of particular measures might be influenced by their specialty and the proportion of their patient populations who were eligible for particular measures; for example, surgeons and anesthesiologists might choose measures related to perioperative surgical care.

^{xviii} Eligible professionals include *Medicare physicians*: Doctor of Medicine, Doctor of Osteopathy, Doctor of Podiatric Medicine, Doctor of Optometry, Doctor of Oral Surgery, Doctor of Dental Medicine, Doctor of Chiropractic; *Practitioners*: Physician Assistant, Nurse Practitioner, Clinical Nurse Specialist, Certified Registered Nurse Anesthetist (and Anesthesiologist Assistant), Certified Nurse Midwife, Clinical Social Worker, Clinical Psychologist, Registered Dietician, Nutrition Professional, Audiologists, Advanced Practice Registered Nurse (APRN); and *Therapists*: Physical Therapist, Occupational Therapist, Qualified Speech/Language Therapist.

Objectives

Two objectives guided the analysis to examine patient and physician characteristics associated with participation in PQRS and the selection of measures PQRS participants choose to report. The objectives for this study are:

1. Determine which physician and patient characteristics are associated with physician participation in the voluntary PQRS.
2. Examine which physician and patient characteristics are associated with the types of measures PQRS participants chose to report.

Methods

To determine which factors affect physician participation, the research team used logistic regression to model two decisions made by physicians. The first decision was whether to participate in PQRS, and the second decision was how to participate, namely, deciding which measures to report. The research team obtained the complete list of physicians submitting claims on Medicare fee-for-service (FFS) beneficiaries and then combined these data with data indicating physician participation in PQRS in 2012, measure reporting choices, and potential explanatory factors from other data sources. The research team then analyzed how physician characteristics, such as specialty, patient case mix, and practice-level factors including practice size and location, affected the two participation decisions. Analysis was restricted to physicians, due to data limitations. In addition, the Meaningful Use and Physician Value-Based Payment Modifier program were excluded because these programs were too recently implemented to allow analyses of participation. Table 3-1 lists the program and setting used in the analyses described in this chapter.

Table 3-1: Program Included in Analysis

Setting	Program	Abbreviation
 Ambulatory	Physician Quality Reporting System	PQRS

Data Sources

Medicare Part B (Part B) physician claims and enrollment data were used to identify physicians who provide services to Medicare beneficiaries and are also eligible to participate in PQRS. These data and Medicare Hierarchical Condition Category (HCC) scores were used to calculate each physician's patient population characteristics and total Medicare FFS payments received. Clinician-level data on measure reporting included in PQRS were used to identify participants as well as to classify the measures reported by participants. Finally, the research team used the Physician Compare database to identify physician and practice characteristics.

Study Population

The research team first identified physicians and their characteristics using the 2012 Part B physician claims and the Physician Compare database. They used 100 percent Part B claims from 2012 to identify individual physicians submitting claims for FFS beneficiaries. Physicians were matched on this list to records in the Physician Compare database via the National Provider Identifier (NPI). The Physician Compare data contain NPI, primary and secondary specialties, clinician gender, years since graduation from medical school, and degree for individual clinicians, as well as practice location and size. A physician practice for which the physician group size was missing was assigned a group size estimated by counting the number of physicians associated with that practice.

The sample was limited to physicians, i.e., those with MD or DO degrees, in Physician Compare. Physicians that were not in Physician Compare were excluded because their individual and practice characteristics were unavailable. Pediatricians were excluded from the analysis because they generally care for few Medicare beneficiaries. Non-physicians, as identified by self-reported degree and specialty, were excluded because such clinicians could not be reliably linked with the Part B claims data, given that these providers may not bill claims under their own NPI. The research team also created 28 specialty categories based on the primary specialty listed for a physician in Physician Compare for use in all analyses: Primary care (which includes internal medicine, family practices, general practice, preventive medicine, and geriatric medicine), anesthesiology, cardiac surgery, cardiology, dermatology, emergency medicine, endocrinology, gastroenterology, general surgery, hematology/oncology, infectious diseases, nephrology, neurology, obstetrics/gynecology, ophthalmology, orthopedic surgery, other medical specialty, other surgical specialty, pain management, pathology, podiatry, psychiatry, pulmonary, radiology, rheumatology, thoracic surgery, urology, and vascular surgery (Appendix 3-1).

CMS collects quality measure data submitted by each participating clinician through claims, a qualified registry, electronic health records (EHRs), or the Group Practice Reporting Option (GPRO) Web interface. The research team used these data to identify physicians participating in PQRS in 2012 and the numbers and types of measures each physician reported. The data were compiled at the physician level and included the PQRS measure identifier, an NPI for the physician, the number of patients submitted as eligible for the measure, and the number of patients passing the measure. The Physician Compare data included information on participation in PQRS as of May 2013 (the date that the data were downloaded), but the research team relied on the PQRS data because these data were more likely to contain accurate information on participation from the relevant year (2012) as compared with the Physician Compare data.

Classifying PQRS Measures

Each measure in PQRS was designed to estimate performance on care related to specific procedures, e.g., coronary artery bypass graft surgery; or care related to primary disease categories, such as diabetes mellitus; preventive care; and care across several domains, such as documentation of medications in the medical record. For the 2012 data, 266 individual measures were collapsed into 60 measure categories to permit tractable analyses. A hyperlink to a list of all measures used in this study is provided in Appendix i-4.³ The list includes the NQF endorsement status, NQF number if endorsed, and both the measure title used by the CMS

program and the measure title used by NQF. The research team based the classification on the title of the measure and the presence of matching International Classification of Diseases 9th edition (ICD-9) codes. For example, measures starting with “Diabetes Mellitus” were classified as “Diabetes Mellitus” measures, if they contained common ICD-9 codes. The research team also assigned each measure to one related specialty for the purposes of visualization. For example, asthma measures were assigned to pulmonary medicine, and measures related to coronary artery disease (CAD) were assigned to cardiologists, with the understanding that other physicians from other specialties may also appropriately report such measures (Appendix i-4).

Adding Patient Population Characteristics

The research team used the Master Beneficiary Summary File, Medicare HCC risk score data, and Part B physician claims data to identify patient clinical and demographic characteristics, to compute an average risk score for each physician’s patients, and to identify the proportion of patients eligible for each of the PQRS measure categories.

The research team constructed a patient-level file, based on the Master Beneficiary Summary File, to identify all Medicare beneficiaries, along with demographic data (age and sex), and then added HCC community risk score data for 2012. The HCC community score is based on beneficiary demographics and healthcare utilization in the prior year; thus, it represents each beneficiary’s baseline status exclusively.⁸ After constructing the patient-level file, the research team linked patients to physicians using the Part B physician claims data. These data include a patient identifier, an NPI, diagnosis (ICD-9), Healthcare Common Procedure Coding System (HCPCS) codes, Current Procedure Terminology (CPT) codes, and the amount paid by Medicare for the claim, i.e., payments to physicians who did not reassign their benefits to a hospital or to a group practice. The research team identified each Medicare beneficiary seen by each physician in each year by first linking the claim to the beneficiary by the Medicare beneficiary identifier on the Part B claim and then linking the claim to the physician by the NPI on the claim. After linking patients to physicians, the research team generated patient population characteristics for each physician. These included the mean patient age, proportion of females, and the mean patient HCC score. They also computed the total annual Part B payments for each physician who did not reassign his/her benefits to a hospital or group practice. Finally, the research team computed the proportion of each physician’s patients eligible for each of the 60 PQRS measure categories. ICD-9 and CPT/HCPCS codes were used to examine each claim’s eligibility for each of the 60 measure categories (Classifying PQRS Measures section). Of note, these data lacked CPT level II codes, which are frequently used in determining whether the patient actually received the care process specified in the measure; however, this analysis primarily concerns eligibility, which is adequately defined by the use of CPT/HCPCS level I codes.

Analysis

Objective 1—Determine Which Physician and Patient Characteristics Are Associated With Physician Participation in the Voluntary PQRS Program

An initial analysis found that PQRS participation varied substantially within practices; therefore, the research team used individual physicians as the unit of analysis. The relationship was modeled between the outcome (an indicator variable that is equal to one for physicians who

participated in PQRS in 2012, and zero, otherwise) and several predictor variables (practice size, census region, urban or rural area, physician specialty, gender, years since medical degree, and total 2012 Part B payments). Because total Part B payments are skewed, with high outlier values, payments were log transformed before estimating models. Log transformation is performed by taking the natural log of the total payments to reduce the influence of high outliers on the estimated relationships.⁹ The research team also included patient population predictors, such as mean age, proportion of women, and average HCC score among the physician's patients.

For each variable, the research team tested for statistically significant differences between participating and non-participating physicians. For continuous variables, e.g., Part B payments, the research team used a t-test or Wilcoxon rank sum test if the variable was not normally distributed. For categorical variables, e.g., practice size category, the research team used a Chi-square test and then performed multivariate logistic regression to estimate the effect of each predictor after controlling for the other predictors simultaneously. The aforementioned analyses tested which factors are associated with physicians adopting particular measures. The research team estimated the change in probability of an outcome, i.e., the “marginal effect,” attributable to a particular predictor, using the “recycled predictions” method.¹⁰ This method calculates the predicted change in the outcome for each observation in the study population due to a change in the predictor. These individual predicted changes are then averaged for the entire study population, yielding the average change (or marginal effect). For example, the change in probability of participating in PQRS attributable to physicians being female can be estimated by taking the difference between the predicted probability of PQRS participation if all physicians were hypothetically female and the predicted probability of participation if all physicians were hypothetically male.

In addition to estimating an effect for the full physician study population, the research team estimated separate models for practice size and specialty categories. The research team also explored possible non-linear relationships between Part B payments and participation to test the hypothesis that physicians might decline to participate unless reimbursements exceeded a threshold level of Part B payments.

Objective 2—Examine Which Physician and Patient Characteristics Are Associated With the Types of Measures PQRS Participants Chose to Report

The research team explored the relationship between measure selection and physician specialty with both graphical and quantitative analyses. The number of measures and specialties were compressed, as shown in Appendix i-4 and Appendix 3-1. The research team selected 16 specialties and 19 measure categories to show in the graphical analysis. The 16 specialties selected were those that submitted the most measure reports to PQRS and had a measure category considered relevant to the specialty, e.g., endocrinology and measures on diabetes mellitus. Also displayed is the proportion of physicians reporting surgical or cross-cutting measures or the health information technology (IT) measure. In the graphical analysis, provider specialty was plotted against measure category. Each row represented the range of potential categories that a specialty could choose to report, and each cell represented the proportion of the measures reported in each category by that specialty. The visualization allowed analysts to see quickly which types of measures are reported by particular specialties.

The research team then used multivariate logistic regression to estimate the relationship between measure selection and various potential predictors. In each model, the study population was

drawn from the population of physicians providing services to Medicare beneficiaries but restricted to PQRS participants. The outcomes of interest are the selection of any measure from one of six highly reported clinical measure categories: Preventive care, which includes routine vaccinations and cancer screenings; diabetes mellitus; stroke; CAD; perioperative care; and urinary incontinence; an additional outcome of interest is the selection of the cross-cutting “Use of Health Information Technology” measure.

The research team modeled the selection of each measure category as a function of PQRS participant characteristics, including practice size, gender, region, average HCC score among patients, and years since medical school. In addition, the research team included the 28 specialty categories previously described in the model: Primary care was the base category to which the other 27 specialties were compared. The research team then performed multivariate logistic regression to estimate the effect of each of the predictors after controlling for every other predictor simultaneously. The analyses tested the importance of physician characteristics on measure selection. The research team again estimated the change in probability of selecting a particular measure category, i.e., the “marginal effect,” attributable to a particular predictor, using the “recycled predictions” method.^{9, 10} The research team calculated the change in probability with respect to the base-case physician (primary care specialty, large practice size, and located in an urban setting in the Western census region).

SAS version 9.4 was used for all statistical analyses. Standard errors were clustered by practice for all statistical tests. The Bonferroni method was used to control for increased Type I error (i.e., cases where an effect would be found where the true effect was zero) associated with multiple comparisons.¹¹

Limitations

First, the research team did not have data on the following variables: EHR usage, the feasibility of measure reporting, how physicians view the measures, how decisions about participation were made by practice leadership, and how bonus payments were shared in larger practices. As a result, the study is unable to assess all potential influences on the decision to participate in PQRS reporting.

Second, Federman and Keyhani found most PQRS participants had little confidence in the program’s impact on quality.⁵ If a disproportionate number of physicians from small practices shared this belief and if this belief was strongly associated with non-participation, then it would be difficult to determine whether non-participation was related to concerns regarding quality or other factors related to larger practice size, such as more financial resources to invest in quality reporting.

Third, another potential influence on the decision to participate is the level of performance on the measures for which a physician is eligible. Physicians may choose to report exclusively on measures where they can demonstrate high performance. Physicians who do not believe they have high performance on any of the measures appropriate to their patient population may choose not to participate. The research team was unable to observe performance on measures that are unreported, so it was unable to analyze the relationship between performance and participation.

Fourth, due to data limitations, this analysis only addressed decisions by physicians. Decision-making may be substantially different among eligible non-physician professionals.

Finally, several analyses use provider specialty, which is self-reported in the Physician Compare database and is not validated. If incorrect values for specialty or coding a secondary specialty as primary are common in the Physician Compare database, then observed associations could be biased. In addition, the research team also relied on provider-level data on quality measures to define PQRS participation and measurement selections; these data are also not validated.

Results

Objective 1—Determine Which Physician and Patient Characteristics Are Associated With Physician Participation in Voluntary PQRS Program

Table 3-2 summarizes the descriptive characteristics of the universe of physicians eligible to participate in PQRS, as defined in the study, i.e., physicians reporting to PQRS who are listed on Physician Compare and who submitted Medicare claims, and the population of physicians participating in PQRS in 2012. For example, 32.5 percent of the physicians eligible to participate in PQRS had a primary specialty that the research team classified as primary care, i.e., general practice, internal medicine, family practice, preventive medicine, and geriatric medicine. Radiology (7.8 percent), anesthesiology (6.4 percent), cardiology (5.6 percent), and obstetrics/gynecology (5.5 percent) have the highest representation among the remaining specialties. The majority of physicians (57.1 percent) were in large practices of 25 or more. Physicians in large practices were more prevalent among PQRS participants than among non-participants. Radiologists, anesthesiologists, and emergency medicine physicians were also substantially more prevalent among PQRS participants than they are among all physicians. PQRS participants have slightly less experience and higher Part B payments than non-participants. PQRS participants were slightly more likely to be located in the South and Midwest and in metropolitan areas than elsewhere. All differences shown in Table 3-2 are statistically significant at the 0.05 level.

Table 3-2: Characteristics of Physicians Included in Study

Variable	Proportion or Mean All MD/DO ^{xix} (n = 650,423)	Participating in PQRS ^{xx} (n = 268,867)	p value ^{iv}
Specialty Category			
Anesthesiology	6.4%	10.7%	<0.0001
Cardiac Surgery	0.3%	0.5%	
Cardiology	5.6%	6.1%	
Dermatology	1.8%	1.4%	
Emergency Medicine	3.0%	4.9%	
Endocrinology	0.9%	1.0%	
Gastroenterology	2.2%	2.1%	
General Surgery	3.5%	2.8%	
Hematology/Oncology	3.0%	3.1%	
Infectious Diseases	0.9%	0.7%	
Nephrology	1.4%	1.2%	
Neurology	2.4%	2.1%	
Obstetrics/Gynecology	5.5%	3.1%	
Ophthalmology	3.1%	4.7%	
Orthopedic Surgery	3.8%	3.0%	
Other	2.0%	1.2%	
Other Surgery	3.8%	2.8%	
Pain Management	0.6%	0.4%	
Psychiatry	4.0%	1.1%	
Pulmonary	2.3%	2.1%	
Radiology	7.8%	12.3%	
Rheumatology	0.7%	0.7%	
Thoracic Surgery	0.4%	0.5%	
Urology	1.7%	2.1%	
Vascular Surgery	0.5%	0.6%	
Primary Care	32.5%	29.1%	
Physician Characteristics			
Female	29.0%	28.2%	<0.0001
Years Since Medical School	21.4 years	20.4 years	<0.0001
Total 2012 Part B Payments	\$117,806	\$136,127	<0.0001

^{xix} Physicians with at least one claim in the 2012 Part B claims file who are MD/DOs identified by training and specialty in the 2012 Physician Compare database

^{xx} Physicians from column 2 reporting at least one measure to PQRS

^{iv} p value for test of differences in participation rates based on chi-square test for categorical variables (specialty category, female, practice size category, region, and rural/urban) and t-test for continuous variables (years since medical school, total Part B payments, percentage of patients female, average patient HCC score)

Table 3-2: Characteristics of Physicians Included in Study

Variable	Proportion or Mean All MD/DO ^{xix} (n = 650,423)	Participating in PQRS ^{xx} (n = 268,867)	p value ^{iv}
Practice Size			
Solo	9.4%	4.2%	<0.0001
2–3 Physicians	7.4%	4.9%	
4–24 Physicians	26.0%	25.1%	
25 or More Physicians	57.1%	65.9%	
Region			
Midwest	24.0%	27.4%	<0.0001
Northeast	21.9%	18.9%	
South	34.4%	36.6%	
West	19.7%	17.2%	
Rural/Urban			
Non-metropolitan	5.5%	4.6%	<0.0001
Metropolitan	94.5%	95.4%	
Patient Population Characteristics			
Percentage of Patients Female	59.3%	58.3%	<0.0001
Percentage of Patients Non-white	31.2%	28.8%	
Average Patient HCC Score	1.59	1.61	

Table 3-3 presents the percentage of physicians that report PQRS measures by practice size and reporting mechanism. Among those who participate, the claims reporting mechanism is the most common method used to report PQRS measures. Reporting via a qualified registry is the second most common method for practices with fewer than 25 physicians. The Group Practice Reporting Option (GPRO) is the second most common data reporting method for practices with 25 or more physicians. Reporting through EHR (both direct submission and data submission vendor) is the least common method.

Table 3-3: PQRS Measure Reporting Mechanism by Practice Size

Practice Size	No Reported Measures	Claims	EHR	Registry	GPRO
Solo	83.2%	15.4%	0.4%	0.9%	0.0%
2–3 Physicians	76.3%	21.8%	0.7%	1.2%	0.0%
4–24 Physicians	65.0%	32.7%	0.7%	1.6%	0.0%
25 or More Physicians	57.2%	31.0%	0.3%	3.3%	8.2%

Table 3-4 shows results from multivariate models assessing factors associated with participation in PQRS. These results are generally consistent with those in Table 3-2. Practice size has the largest estimated association with participation. Compared with physicians in large practices, those in solo practices have a 29.7 percentage point lower likelihood of participation in PQRS. Participation rates varied significantly among specialties. Anesthesiologists have a statistically significant 27.7 percentage point higher likelihood of participation than primary care physicians. Physicians whose primary specialty was emergency medicine (+26.8 percentage points),

ophthalmology (+18.7 percentage points), and radiology (+18.7 percentage points) also had much higher likelihoods of participation than primary care physicians. Alternatively, psychiatrists had a 23.0 percentage point lower likelihood of participation. Female gender was associated with a small increase in probability of participation, in contrast to results derived from univariate analyses (Table 3-2). Practice location and percentage of female patients were not statistically significant in the multivariate model, after controlling for multiple testing. While Part B payments for physicians who did not reassign their benefits had a statistically significant association with PQRS participation, the magnitude of the relationship is small: An increase in annual payments of \$10,000 over the average is associated with an increase of 0.5 percentage points in the likelihood to participate. Physicians with sicker populations, i.e., higher HCC scores, were less likely to participate in PQRS (-3.3 percentage points for a one unit change); this difference translates into a 4.2 percentage points lower participation rate in practices whose average HCC score was at the 90th percentile, as compared with practices whose average HCC score was at the 10th percentile, with a lower percentile indicating a healthier population. Physicians with a higher proportion of non-white patients had lower participation rates: -12.0 percentage points for a one unit change. This difference corresponds to a 5.9 percentage point lower participation rate in practices whose average non-white proportion was at the 90th percentile, as compared with practices whose average non-white proportion was at the 10th percentile.

Table 3-4: Factors Affecting Participation (Multivariate Model Results)

Variable	Effect on Probability of PQRS Participation ^{xxi}	<i>p</i> value ^{xxii}
Specialty Category		
Anesthesiology	27.7%	<0.0001
Cardiac Surgery	16.0%	<0.0001
Cardiology	-1.6%	0.102
Dermatology	-6.8%	<0.0001
Emergency Medicine	26.8%	<0.0001
Endocrinology	8.2%	<0.0001
Gastroenterology	0.3%	0.81
General Surgery	-4.1%	<0.0001
Hematology/Oncology	-3.5%	0.033
Infectious Diseases	-2.5%	0.066
Nephrology	-4.1%	0.006
Neurology	-2.4%	0.040
Obstetrics/Gynecology	-5.8%	<0.0001
Ophthalmology	18.7%	<0.0001
Orthopedic Surgery	-9.3%	<0.0001
Other	-11.4%	<0.0001
Other Surgery	-6.2%	<0.0001

^{xxi} From multivariate logistic regression with participation in PQRS as the outcome and variables in column one as predictors; effects are estimated using recycled prediction, i.e., by comparing probability of participation in one group versus the “base” group. (See Methods section for additional details.)

^{xxii} *p* values for the coefficient estimates from multivariate logistic regression with participation in PQRS as the outcome and variables in column one as predictors. Bonferroni adjusted 0.05 statistical significance level is 0.0005.

Table 3-4: Factors Affecting Participation (Multivariate Model Results)

Variable	Effect on Probability of PQRS Participation ^{xxi}	p value ^{xxii}
Pain Management	-10.2%	<0.0001
Psychiatry	-23.0%	<0.0001
Pulmonary	-1.5%	0.122
Radiology	18.7%	<0.0001
Rheumatology	-1.3%	0.39
Thoracic Surgery	10.9%	<0.0001
Urology	8.6%	<0.0001
Vascular Surgery	3.3%	0.018
Primary Care	Base	
Physician Characteristics		
Female	1.8%	<0.0001
Years Since Medical School	-0.3%	<0.0001
Total 2012 Part B Payments (log)	7.4%	<0.0001
Practice Size		
Solo	-29.7%	<0.0001
2–3 Physicians	-22.8%	<0.0001
4–24 Physicians	-11.4%	<0.0001
25 or More Physicians	Base	
Region		
Midwest	6.2%	0.005
Northeast	-4.2%	0.030
South	3.7%	0.034
West	Base	
Rural/Urban		
Non-metropolitan	-3.8%	0.033
Metropolitan	Base	
Patient Population Characteristics		
Percentage of Patients Female	-4.8%	0.007
Percentage of Patients Non-white	-12.0%	<0.0001
Average Patient HCC Score	-4.1%	<0.0001

The research team explored separate models for each practice size category to determine if variables had different effects in each category, as shown in Table 3-5. The association between specialty and participation varies in magnitude by practice size but is consistent with the association observed in the model that uses data from all practice sizes. For each increase of \$10,000 above the average in total payments, the likelihood of participation increased by 0.4 percentage points for physicians in solo practices and 0.5 percentage points for those in practices with 25 or more. The estimated association between the proportion of non-white patients and participation was larger for physicians in larger practices. Among physicians in solo practices, the participation rate in practices whose average non-white proportion was at the 90th percentile was 1.6 percentage points lower than practices whose average non-white proportion was at the

10th percentile. For physicians in practices with 25 or more physicians, the likelihood was 7.1 percentage points lower.

Table 3-5: PQRS Participation by Practice Size

Variable ^{xxiii}	Solo Practice ^{xxiv}		2–3 Physicians		4–24 Physicians		25+ Physicians	
Specialty Category								
Anesthesiology	8.7%	**	3.9%		29.0%	**	29.4%	**
Cardiac Surgery	9.1%		20.4%	**	19.2%	**	14.0%	**
Cardiology	-5.2%	**	-3.9%		-2.7%		-0.2%	
Dermatology	-4.4%	**	-5.2%	**	-9.4%	**	-4.6%	
Emergency Medicine	26.8%	**	33.9%	**	34.3%	**	19.0%	**
Endocrinology	1.2%		-3.7%		4.9%		11.4%	**
Gastroenterology	-1.9%		-5.8%		-1.0%		2.8%	
General Surgery	-6.9%	**	-6.9%	**	-3.9%		-3.0%	
Hematology/Oncology	-1.9%		0.6%		1.0%		-6.1%	
Infectious Diseases	-12.5%	**	-17.9%	**	-12.7%	**	4.2%	
Nephrology	-2.3%		-6.1%	**	-13.9%	**	1.2%	
Neurology	-6.0%	**	-9.3%	**	-8.4%	**	1.4%	
Obstetrics/Gynecology	-7.9%	**	-8.1%	**	-9.1%	**	-2.9%	
Ophthalmology	14.2%	**	23.2%	**	25.3%	**	8.5%	
Orthopedic Surgery	-8.9%	**	-11.1%	**	-9.8%	**	-7.5%	**
Other	-11.8%	**	-15.9%	**	-18.2%	**	-4.7%	
Other Surgery	-10.5%	**	-12.8%	**	-9.6%	**	-0.5%	
Pain Management	-7.4%	**	-11.6%	**	-14.6%	**	-3.9%	
Psychiatry	-16.4%	**	-22.4%	**	-30.6%	**	-20.4%	**
Pulmonary	-0.9%		-3.8%		-4.1%		-0.4%	
Radiology	4.8%		9.5%	**	26.3%	**	15.7%	**
Rheumatology	-2.9%		-4.9%		-7.1%		2.6%	
Thoracic Surgery	3.3%		15.9%		15.0%	**	10.2%	**
Urology	5.7%	**	5.3%		11.7%	**	7.8%	**
Vascular Surgery	-4.9%		-6.1%		6.8%		4.4%	
Primary Care	Base		Base		Base		Base	
Physician Characteristics								
Female	3.6%	**	2.5%	**	0.7%		0.7%	
Years Since Medical School	-0.4%	**	-0.3%	**	-0.2%	**	-0.1%	**
Total 2012 Part B Payments (log)	5.8%	**	4.5%	**	6.1%	**	7.1%	**

^{xxiii} The numbers reported in the table are the estimated marginal effects from a multivariate logistic regression with participation in PQRS as the outcome and variables in column 1 as predictors. There is one model for each practice size category.

Marginal effects are estimated using recycled prediction. (See Methods section for additional details.)

^{xxiv} **Statistically significant at the 0.05 level or below after adjusting for multiple testing.

Table 3-5: PQRS Participation by Practice Size

Variable ^{xxiii}	Solo Practice ^{xxiv}		2–3 Physicians		4–24 Physicians		25+ Physicians	
Region								
Midwest	3.6%	**	4.1%	**	3.8%	**	7.1%	
Northeast	0.1%		1.2%		-2.1%		-5.9%	
South	3.0%	**	3.9%	**	2.9%		3.5%	
West	Base		Base		Base		Base	
Rural/Urban								
Non-metropolitan	-0.5%		-2.0%		-7.5%	**	-3.3%	
Metropolitan	Base		Base		Base		Base	
Patient Population Characteristics								
Percentage of Patients Female	1.4%		0.7%		2.4%		5.3%	
Percentage of Patients Non-white	-3.4%	**	-10.3%	**	-13.1%	**	-14.4%	
Average Patient HCC Score	-2.7%	**	0.3%		2.1%		-5.1%	**

Figure 3-1 shows the relationship between Part B payments and participation in PQRS by practice size. The x-axis is the percentile of total Part B payments for all physicians in 2012 divided into 20 bins, each representing a 5-percentile range. The y-axis is the participation rate for physicians in each payment bin. The participation rate for physicians in solo practices who have total Part B payments in the lowest bin is 2 percent. This rate increases to 27 percent for physicians in solo practices who have total payments in the highest bin. In contrast, the participation rate for physicians in practices with 25 or more is 14 percent for the lowest payment bin and 57 percent for the highest. The highest participation rate for physicians in solo practices at any percentile remains below 30 percent, even for those who have the total payments in the top 5 percent, while physicians in the largest practices who have total payments as low as the 15th percentile have a 36 percent participation rate.

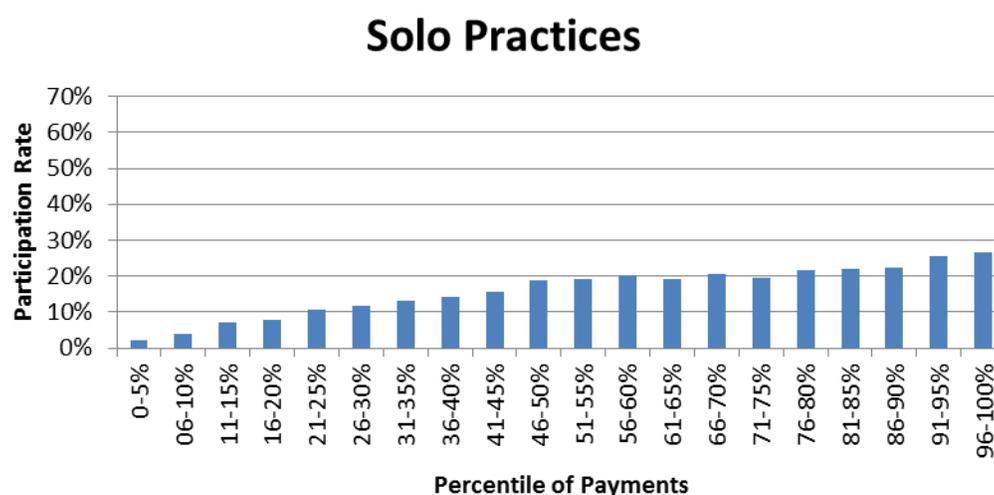
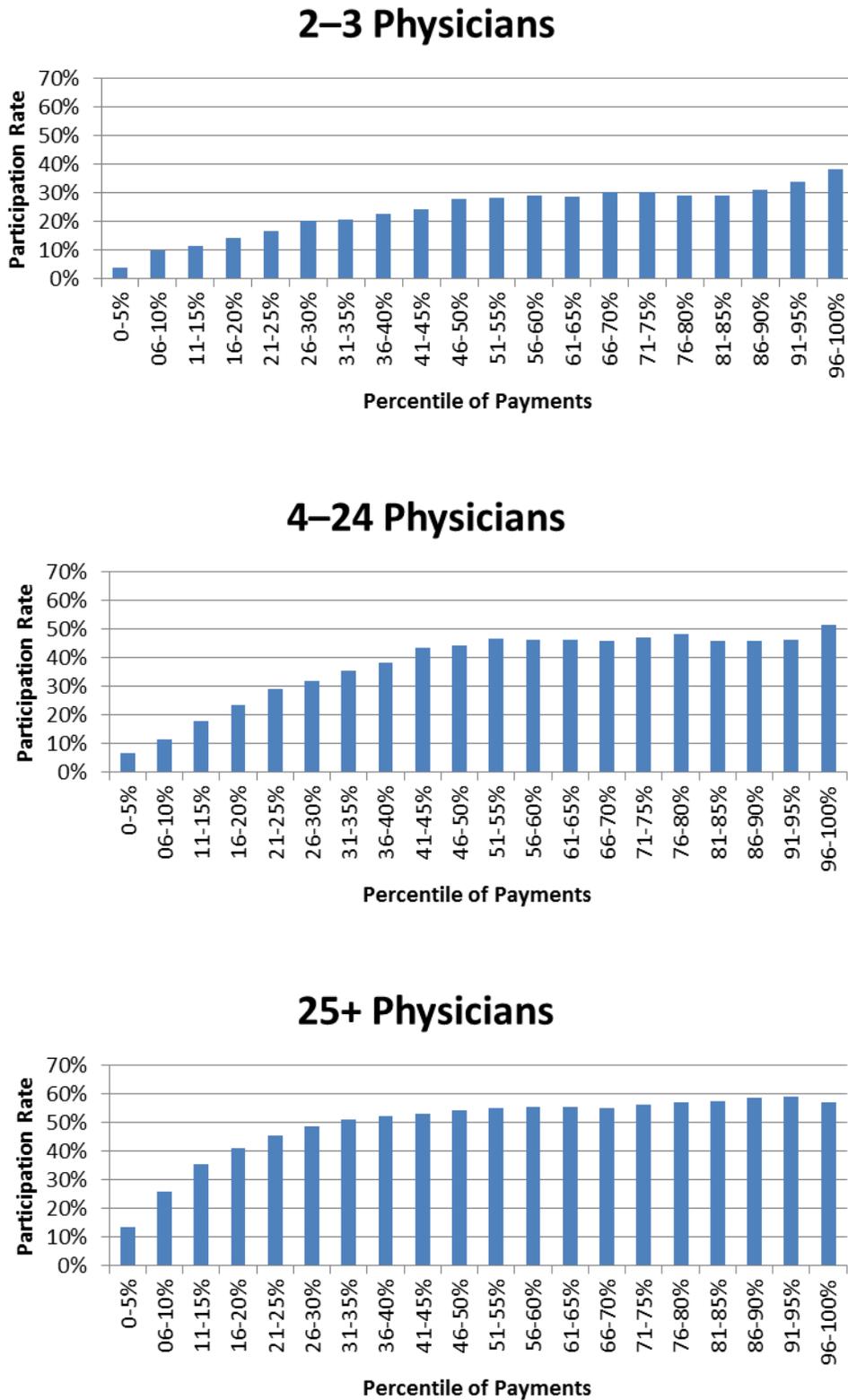
Figure 3-1: PQRS Participation Rate by Part B Payments

Figure 3-1: PQRS Participation Rate by Part B Payments



Objective 2—Examine Which Physician and Patient Characteristics Are Associated With the Types of Measures PQRS Participants Chose to Report

Of the 266 measures available in 2012, PQRS participants (including non-physicians) reported 219. The graphical analysis exploring the relationship between physician specialty and the reported measure category for selected measures is shown in Figure 3-2. Each row represents the range of potential categories that a specialty could choose to report, and each cell represents the proportion of the measures reported in each category by that specialty. Darker shades correspond to higher proportions. The pattern in Figure 3-2 suggests that the majority of physicians report measures relevant to their specialty. Most of the “out-of-specialty” selections occurred because physicians selected primary care and cross-cutting measures. For example, as shown by varying shades of the top row of the Figure 3-2, 14.1 percent of measures reported by cardiologists are cardiology measures (the darkest cell), 33.3 percent are cross-cutting measures, and 27.3 percent are primary care measures, while 17.7 percent are endocrinology measures. Most specialties chose to submit primary care, cross-cutting, and endocrinology (generally diabetes) measures. These measures make up a high proportion of reported measures for all specialties in Figure 3-2 except emergency medicine, pathology, ophthalmology, and radiology. The latter four specialties have the highest proportion of reported measures corresponding to their specialty. In addition, anesthesiologists have primarily submitted surgical measures (>90 percent not shown), which are clearly relevant to their practice.

Figure 3-2: Proportion of PQRS Measure Categories Reported by Each Specialty

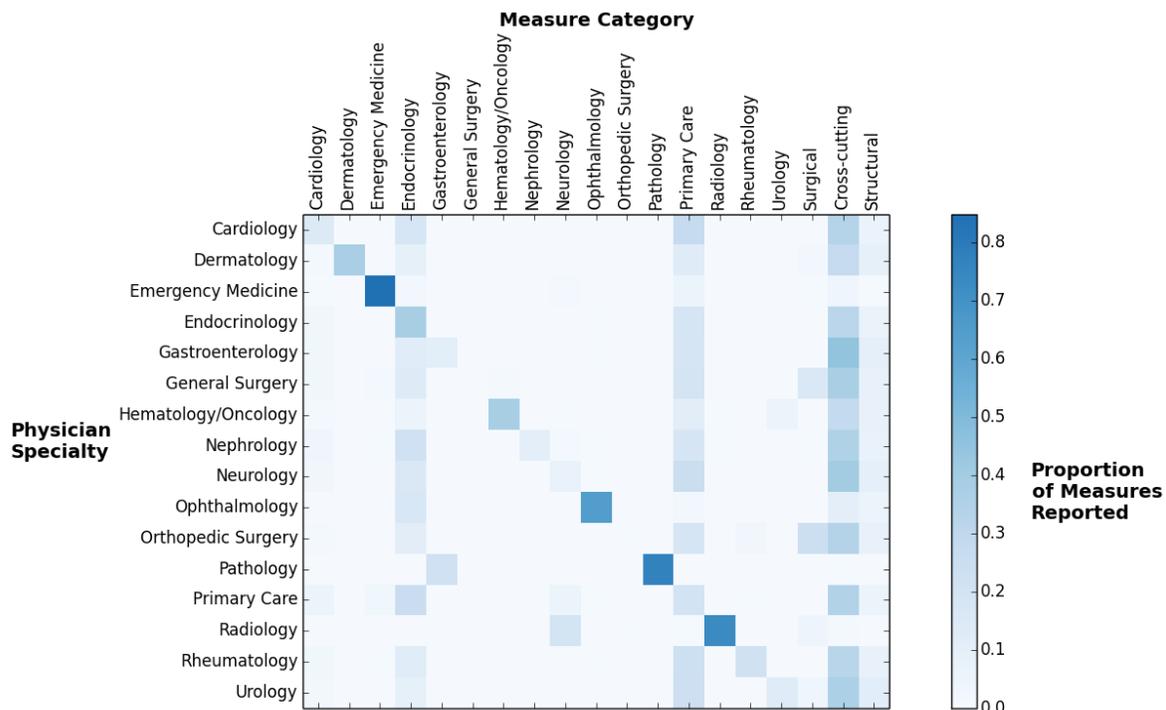
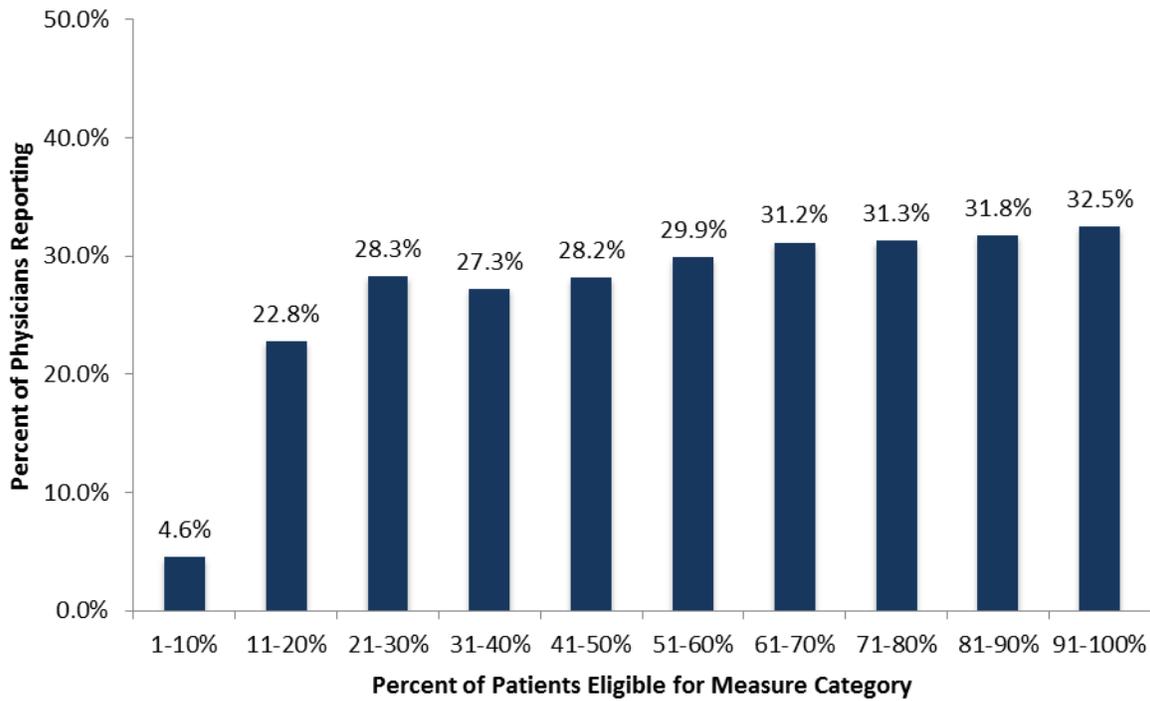


Figure 3-3 shows the relationship between patient disease profile and disease-specific measure reporting. Physicians participating in PQRS are divided into groups, based on the percentage of their patients who are eligible for any of the PQRS measure categories. Figure 3-3 shows the percentage of physicians in each group who report at least one measure in the measure category. A positive relationship exists between the proportion of patients eligible for a measure category and the reporting rate. Of the physicians who have 1 percent to 10 percent of their patients eligible for a measure category, 4.6 percent report at least one measure within that category; of those with 91 percent to 100 percent of their patients eligible for a category, 32.5 percent report at least one measure within that category.

Figure 3-3: Measure Reporting Rates by Percentage of Patients Eligible for Measure Category



Multivariate Analyses

Figure 3-2 and Figure 3-3 suggest that clinician specialty and measure choice are correlated, as might be patient population and measure choice. The research team analyzed specialty and patient population together to determine whether there are differences in how they each influence measure selection. Each regression result computes the marginal effect of particular specialties relative to primary care.

Table 3-6 summarizes three additional categories of greater relevance to other specialties: Use of health IT, perioperative care, and urinary incontinence. Measure selection choices were consistent with expectations, with those involved in surgical care, including anesthesiologists (+70.2 percent), cardiac surgeons (+30.5 percent), orthopedic surgeons (+43.3 percent), and general surgeons (+29.5 percent), likely to select measures related to perioperative care. Urologists were more likely to report measures related to urinary incontinence (+22.2 percent compared with primary care physicians), but other surgical specialties were less likely than primary care physicians to report such measures. Use

of health IT reporting was higher in primary care physicians in comparison with anesthesiologists (-30.1 percent), emergency medicine specialists (-22.7 percent), and radiologists (-31.8 percent).

Table 3-6: Multivariate Models of Measure Selection – 1^{xxv}

Predictor	HIT ^{xxvi}		Perioperative Care		Urinary Incontinence	
Specialty = Anesthesiology	-30.1%	**	70.2%	**	-12.2%	**
Specialty = Cardiac Surgery	-8.6%	**	30.5%	**	-6.6%	**
Specialty = Cardiology	18.2%	**	6.4%	**	6.1%	**
Specialty = Dermatology	-7.3%	**	-15.7%	**	-8.6%	**
Specialty = Emergency Medicine	-22.7%	**	1.6%		-9.5%	**
Specialty = Endocrinology	4.3%		-2.6%		-3.3%	**
Specialty = Gastroenterology	20.1%	**	-4.8%		-3.6%	**
Specialty = General Surgery	7.1%	**	29.5%	**	-1.8%	
Specialty = Hematology/Oncology	-1.1%		-5.8%		-6.4%	**
Specialty = Infectious Diseases	16.5%	**	-15.8%	**	3.1%	
Specialty = Nephrology	12.9%	**	-0.9%		4.2%	
Specialty = Neurology	13.3%	**	1.7%		-1.7%	
Specialty = Obstetrics/Gynecology	11.0%	**	16.1%		6.0%	**
Specialty = Ophthalmology	-11.7%	**	-3.9%		-11.0%	**
Specialty = Orthopedic Surgery	-0.6%		43.3%	**	-6.2%	**
Specialty = Other	12.6%	**	24.6%	**	-4.8%	**
Specialty = Other Surgery	9.5%	**	28.0%	**	-6.1%	**
Specialty = Psychiatry	12.8%		-15.7%	**	-4.7%	
Specialty = Pulmonary	11.2%	**	10.6%		1.8%	
Specialty = Radiology	-31.8%	**	7.7%		-12.1%	**
Specialty = Rheumatology	4.6%		-1.9%	**	-4.2%	**
Specialty = Thoracic Surgery	-2.6%		34.3%	**	-5.5%	
Specialty = Urology	10.6%	**	27.5%	**	22.2%	**
Specialty = Vascular Surgery	5.0%		38.8%	**	-3.2%	
Specialty = Primary Care	Base		Base		Base	
Practice Size = Solo Practice	-3.0%		-0.7%		0.1%	
Practice Size = 2–3 Physicians	-4.2%		-0.5%		-0.4%	
Practice Size = 4–24 Physicians	-5.0%	**	0.5%		-1.1%	
Practice Size = 25+ Physicians	Base		Base		Base	
Years Since Medical School	1.1%	**	1.2%		-7.9%	
% of Patients Eligible for Measure	-0.08%	**	0.0003%	**	0.02%	**
Average Patient HCC Score	0.3%		0.2%		0.1%	**

^{xxv} The numbers reported in the table are the estimated marginal effects from a multivariate logistic regression with reporting at least one measure in the category as the outcome and variables in column 1 as predictors. There is one model for each measure category. Marginal effects are estimated using recycled prediction and are reported in relation to the base-case physician (specialization in primary care, practicing in a practice with 25+ physicians). (See Methods section for additional details.)

^{xxvi} **Statistically significant at the 0.05 level or below after adjusting for multiple testing.

The physician's patient population was associated with statistically significant increases in measure selection, but the effects were small in magnitude after accounting for specialty. For example, increasing the proportion of a physician's patients eligible for a particular measure from 10 percent to 50 percent was predicted to increase the probability of selecting these measures by less than 2 percent. Practice size influenced measure selection in certain cases. Solo practitioners were less likely to choose to report primary care measures (-6.7 percent probability for reporting preventive care measures compared with physicians from practices with 25 or more physicians).

Table 3-7 summarizes the multivariate analyses of four highly reported measure categories that appeared to be of particular relevance to primary care providers, neurologists, and other internal medicine subspecialists. These categories, each represented by a column in Table 3-7, are preventive care, which includes routine vaccinations and cancer screenings; diabetes mellitus; stroke; and CAD. Specialty was strongly associated with measure selection in these areas, and the magnitudes and directions generally confirmed the graphical results in Figure 3-2. For example, compared with primary care physicians, cardiologists were much more likely to report measures related to CAD (+20.5 percentage points, as shown in the last column of the third row). A cardiac surgeon is estimated to be 18.4 percentage points less likely to report a Preventive Care measure than a primary care physician (as shown in the second row). Compared with primary care physicians, surgeons, anesthesiologists, and dermatologists were less likely to report preventive care measures, although specialties that also frequently provide preventive care, including physicians specializing in obstetrics and gynecology, nephrology, and neurology, were more likely to report preventive care measures. Contrary to expectations, having a primary specialty of neurology was associated with a decreased probability of selecting a stroke measure, confirming the visual results that showed neurologists selecting structural measures, such as health IT reporting, or primary care measures.

Table 3-7: Multivariate Models of Measure Selection – 2^{xxvii}

Predictor ^{xxviii}	Preventive Care		Diabetes Mellitus		Stroke		CAD	
Specialty = Anesthesiology	-42.5%	**	-30.1%	**	-16.4%	**	-13.0%	**
Specialty = Cardiac Surgery	-18.4%	**	-25.2%	**	-14.9%	**	-6.5%	**
Specialty = Cardiology	16.7%	**	-10.6%	**	-16.4%	**	20.5%	**
Specialty = Dermatology	-22.9%	**	-21.3%	**	-15.1%	**	-8.7%	
Specialty = Emergency Medicine	-29.9%	**	-21.8%	**	-7.7%	**	-10.2%	**
Specialty = Endocrinology	-4.8%		-1.5%		-11.2%	**	-6.3%	**
Specialty = Gastroenterology	21.4%	**	-11.4%	**	-14.3%	**	-3.3%	
Specialty = General Surgery	2.2%		-11.9%	**	-13.8%	**	-3.3%	
Specialty = Hematology/Oncology	-8.6%	**	-16.3%	**	-13.2%	**	-7.4%	
Specialty = Infectious Diseases	20.6%	**	9.5%	**	-13.2%	**	-0.8%	
Specialty = Nephrology	13.4%	**	19.0%	**	-13.5%	**	3.2%	

^{xxvii} The numbers reported in the table are the estimated marginal effects from a multivariate logistic regression with reporting at least one measure in the category as the outcome and variables in column 1 as predictors. There is one model for each measure category. Marginal effects are estimated using recycled prediction. (See Methods section for additional details.)

^{xxviii} **Statistically significant at the 0.05 level or below after adjusting for multiple testing.

Table 3-7: Multivariate Models of Measure Selection – 2^{xxvii}

Predictor ^{xxviii}	Preventive Care		Diabetes Mellitus		Stroke		CAD	
Specialty = Neurology	6.3%	**	-10.6%	**	-7.0%	**	-4.3%	
Specialty = Obstetrics/Gynecology	7.0%	**	-16.2%	**	-15.1%	**	-7.4%	**
Specialty = Ophthalmology	-23.6%	**	9.8%	**	-15.1%	**	-11.9%	**
Specialty = Orthopedic Surgery	-12.1%	**	-16.7%	**	-14.8%	**	-8.4%	**
Specialty = Other	1.0%		-14.2%	**	-9.6%	**	-6.7%	**
Specialty = Other Surgery	-1.4%		-16.5%	**	-13.4%	**	-7.2%	
Specialty = Psychiatry	-0.3%		-16.7%		-13.6%		-8.4%	
Specialty = Pulmonary	17.4%	**	-5.5%	**	-9.4%	**	-0.3%	
Specialty = Radiology	-42.0%	**	-30.4%	**	27.8%	**	-13.0%	**
Specialty = Rheumatology	-6.3%		-10.7%	**	-13.9%	**	-3.0%	
Specialty = Thoracic Surgery	-12.3%	**	-23.9%	**	-14.4%	**	-2.8%	
Specialty = Urology	-4.6%		-16.3%	**	-15.0%		-6.0%	
Specialty = Vascular Surgery	-1.4%		-17.7%	**	-14.9%	**	0.2%	
Specialty = Primary Care	Base							
Practice Size = Solo Practice	-6.7%	**	-4.6%	**	-5.0%	**	-3.4%	**
Practice Size = 2–3 Physicians	-5.7%	**	-3.7%	**	-4.6%	**	-2.6%	
Practice Size = 4–24 Physicians	-4.8%	**	-4.4%	**	0.6%		-1.6%	
Practice Size = 25+ Physicians	Base							
Years Since Medical School	-6.3%		-15.8%	**	5.5%	**	-6.7%	
% of Patients Eligible for Measure	0.03%	**	0.05%	**	-0.15%	**	-0.001%	**
Average Patient HCC Score	0.2%	**	0.5%	**	0.4%	**	0.2%	**

Discussion

The research team identified several key variables, including practice size, average patient HCC score, and proportion of non-white patients, associated with decisions by physicians to participate in PQRS. They also found specialty to be a key influence on measure selections among participants; as expected, most participating physicians chose to report measures relevant to their specialty. After adjusting for specialty, patient population characteristics did not appear to have major effects on selection decisions. These findings add to prior work by quantifying the substantial role that practice size plays in explaining PQRS participation. The research team also found that physicians with sicker patient populations and physicians with a higher proportion of non-white patients were somewhat less likely to participate in PQRS.

Given the analytic framework, these results represent associations between various physician decisions and characteristics, but do not necessarily represent causal relationships. Nevertheless, practice size may influence PQRS participation because, in comparison with smaller organizations, large organizations may be able to reduce average costs by spreading the fixed costs associated with participation over a larger group of physicians, i.e., by hiring administrative staff dedicated to collecting and reporting the quality data. This would help explain why relatively small financial incentives (less than 2 percent of total payments) may not have had large effects on inducing participation among physicians, particularly those in smaller practices.³

Indeed, within a given practice size category, total payments, which are correlated with PQRS bonus payments, were not associated with substantially higher participation. The variation in participation rates by specialty may be related to differences in availability of measures relevant to specialty or may be due to differences in setting. For example, radiologists and anesthesiologists, who have higher participation rates, may be more likely to practice in settings, like hospitals, with more institutional resources to assist individual physicians with participation. Lower participation rates among physicians with higher proportions of non-white patients and sicker patients may be due to such physicians having limited resources, and in turn, less infrastructure, e.g., data systems and staffing, to support quality measurement and reporting. Physicians who practice in safety net clinics, for example, may see sicker patients and have less time or resources available to participate in PQRS. Additional research to understand factors that hinder participation could be useful toward developing policies and supporting tools that could enhance physician participation.¹²

These findings extend earlier work on PQRS participation, which focused on the association between individual factors and PQRS participation.³ In this analysis, the research team limited the sample to physicians providing services to Medicare FFS beneficiaries who also report information on Physician Compare. Although this reduced the sample by 40 percent, by using a multivariate approach with these data, the research team was able to identify practice size and healthier patient populations as potentially key factors that might influence PQRS participation decisions.

Regarding measure selection, physicians appear to choose measures that reflect their primary self-reported specialty. While choice of measure selection was not addressed in prior studies, specialty societies have provided recommendations on which measures to report.¹³ Such recommendations may have influenced measure selection decisions, as have CMS regulations requiring a minimum number of patients in the denominator to report measures.

Conclusions

Key Findings

Several key variables were found to be associated with decisions by physicians to participate in PQRS. These factors include practice size, total Part B payments received, average patient HCC score, and proportion of non-white patients. Physician participation in PQRS in 2012 was strongly associated with large practice size and weakly associated with total Part B payments, having a healthier patient population and fewer non-white patients.

Participants were likely to choose measures that were relevant to their practice specialty and those that reflected their patient population. Measure selection was more strongly associated with physician specialty than the proportion of patients eligible for the measures.

Actions to Consider

Participation in PQRS is increasing, but not yet universal. However, recent downward payment adjustments (starting in 2013) and the implementation of the physician value-based payment modifier may significantly increase participation by physicians. In addition, advancements in EHRs and increased alignment of measure specifications, reporting periods, and data sources across quality measurement programs could further reduce the perceived barriers to participation among specialties with lower participation rates.

Action to Consider: Strengthen efforts to make quality data collection easier and less expensive for physician practices by emphasizing EHR-based reporting and measure alignment between quality measurement programs.

Physicians in small practices, physicians with less healthy patients, and physicians with greater proportions of non-white patients appeared to be less likely to participate in PQRS.

Action to Consider: Determine what barriers exist regarding physician participation in quality reporting programs. Consider conducting a survey of non-participating eligible professionals to ascertain what barriers to participation may exist and what actions could be taken to support eligible professionals in reporting of quality measures.

Action to Consider: Consider whether and to what extent the Quality Innovation Network-Quality Improvement Organizations (QIN-QIOs) under the 11th Scope of Work (SOW) can be directed to provide more focused technical assistance to the under-represented practices.

Reference List

- (1) Tax Relief and Health Care Act, §101(b), Pub. L. 109-423; 120 Stat 2975, US Congress (2006).
- (2) Centers for Medicare & Medicaid Services. *2011 Physician Quality Reporting System (PQRS) and Electronic Prescribing (eRx) Incentive Program Experience Report*. Baltimore, MD: Centers for Medicare & Medicaid Services; April 19, 2013.
- (3) Centers for Medicare & Medicaid Services. *2012 Reporting Experience Including Trends (2007-2013): Physician Quality Reporting System and Electronic Prescribing (eRx) Incentive Program*. Baltimore, MD: Centers for Medicare & Medicaid Services; March 14, 2014.
- (4) Laschober M, Maxfield M, Felt-Lisk S, Miranda DJ. Hospital response to public reporting of quality indicators. *Health Care Financ Rev*. 2007;28(3):61-76.
- (5) Federman AD, Keyhani S. Physicians' participation in the Physicians' Quality Reporting Initiative and their perceptions of its impact on quality of care. *Health Policy*. 2011;102(2-3):229-234.
- (6) Hearld LR, Alexander JA, Shi Y, Casalino LP. Pay-for-performance and public reporting program participation and administrative challenges among small- and medium-sized physician practices. *Med Care Res Rev*. 2014;71(3):299-312.
- (7) Berman B, Pracilio VP, Crawford A, et al. Implementing the physician quality reporting system in an academic multispecialty group practice: lessons learned and policy implications. *Am J Med Qual*. 2013;28(6):464-471.
- (8) Li P, Kim MM, Doshi JA. Comparison of the performance of the CMS Hierarchical Condition Category (CMS-HCC) risk adjuster with the Charlson and Elixhauser comorbidity measures in predicting mortality. *BMC Health Serv Res*. 2010;10:245.
- (9) Greene WH. *Econometric Analysis*. 3rd ed. Upper Saddle River, NJ: Prentice Hall; 1997. 339-350.
- (10) Kleinman LC, Norton EC. What's the Risk? A simple approach for estimating adjusted risk measures from nonlinear models including logistic regression. *Health Serv Res*. 2009;44(1):288-302.
- (11) Dunn OJ. Multiple Comparisons among Means. *Journal of the American Statistical Association*. 1961;56(293):52-64.
- (12) Chin MH. Quality improvement implementation and disparities: the case of the health disparities collaboratives. *Med Care*. 2011;49 Suppl:S65-S71.
- (13) American College of Chest Physicians. Guidelines and Resources: Payment Practice and Quality—PQRS. *American College of Chest Physicians*. 2014. Available at: <http://www.chestnet.org/Guidelines-and-Resources/Payment-Practice-and-Quality/Practice-Resources/PQRS-Resources>. Accessed July 2, 2014.

Chapter 4—Measure Alignment: CMS, State, and Veterans Health Administration Measures



Questions on Adoption

1. What measures do the states and the Veterans Health Administration (VHA) adopt?
2. How well are the measures used by the states and the VHA aligned with measures used by the Centers for Medicare & Medicaid Services?

Abstract

Overview: Healthcare entities implement measures to demonstrate accountability and improve the delivery of high-quality care. However, there is variation or lack of alignment in how the measure implementers define measure data elements such as age, target population, and reporting of measure rates. Lack of alignment in measures has the potential to diminish measure impact by creating confusion or inaccurate interpretations of healthcare quality. This study evaluates alignment of measures used by states, the Veterans Health Administration (VHA), and the Centers for Medicare & Medicaid Services (CMS). This study builds on previous alignment studies of measures in public and private health sectors and addresses a knowledge gap regarding alignment of state and federal measures.

Methods: The research team used frequency analysis to assess the alignment of the measures used by states and the VHA with measures used by CMS. Alignment is defined as measures addressing the same measure concept and having the same definitions for age range, target population, measurement period, data source, code sets, and other identifiable measure characteristics. Five alignment categories were used in this study: (1) aligned (used by CMS and consistent with specifications used by CMS), (2) similar (same concept as a measure used by CMS with differences in specifications), (3) not aligned or similar and developed by the state or the VHA and not used by CMS, (4) not aligned or similar and developed by a known developer other than the state or the VHA and not used by CMS (other-standard), and (5) undetermined. The research team explored alignment by evaluating the latest available measure information obtained from state programs and the VHA between

November 1, 2013, and April 30, 2014, and compared the information to measure specifications used by CMS as of December 31, 2013. To identify these measures, the research team reviewed VHA and state websites and contacted state and VHA staff to obtain health quality measures, relying primarily on information contained in state websites. The research team identified 7,005 measures used by states and 514 measures used by the VHA. State health officials did not validate the list of measures collected.

Results: Alignment of measures used by states and CMS varied depending on the state and the program. The percentage of aligned measures (same concept with same specifications) used in state Medicaid programs (45 percent) and state hospital report cards (47 percent) was higher compared to the percentage of similar measures (same concept with different specifications) in these programs (13 percent and 3 percent, respectively). In the state healthcare-associated infection reporting programs, the percentage of similar measures (45 percent) was higher than aligned measures (33 percent). The state nursing home pay-for-performance programs have the highest percentage of locally developed measures (59 percent) as compared to other state programs. In the VHA measurement program, the percentage of aligned measures was 19 percent and the percentage of similar measures was 16 percent. Analyses of measures used by the VHA also showed a high percentage of measures developed by the VHA (65 percent).

Conclusions: The high percentage of locally developed measures in state and VHA programs may indicate a lack of available measures to address priority measurement areas and reflect differences in populations served by these entities. There is a need to further understand the underlying reasons for the variation and lack of measure alignment between programs and to explore the potential impact of alignment issues on consumers and providers.

Background

Measure alignment, or the implementation of the same measures with the same measure specifications, permits reliable assessment of care and accelerates the drive toward a higher standard of quality care.^{6,7} In the past decade, there has been a proliferation of quality measures, and they are increasingly being used for accountability purposes in the private and public healthcare sectors. However, previous studies have identified that public and private organizations that implement the same measure concepts but use different measurement specifications and have varying reporting requirements that result in lack of alignment in the use of quality measures.^{1,2,3,4,5} Thus, as quality measures have become an established feature of the healthcare landscape, the movement toward aligning and reducing variability in measures is a critical step toward providing efficient and effective measurements for diverse populations. This study contributes to current measure alignment knowledge by including other setting-specific measures not included in previous studies and will help inform current and future measure alignment efforts by providing a descriptive analysis of the degree of alignment between the

Centers for Medicare & Medicaid Services (CMS), Veterans Health Administration (VHA), and state-used measures.

One element of the CMS Quality Strategy mission is to “lead quality measure alignment.”⁸ To act on this mission, it is important to understand how measures are aligned among various entities implementing quality measures. CMS and other entities, including state Medicaid agencies and other state and federal governmental partners, can collaboratively shape and implement initiatives to advance measure alignment.

There are compelling reasons for evaluating the existing degree of alignment among measures used by CMS, states, and the VHA. Providers participating in state quality measurement programs such as Medicaid may also be required to report on measures used in CMS quality measurement programs. Likewise, there is overlap in the measures used by the VHA and CMS. For example, hospital measures used by the VHA appear on the CMS *Hospital Compare* website and the *VA Compare* website. Thus, an evaluation of measure alignment among these three entities may help reveal measurement inconsistencies that contribute to provider burden and inability to gauge provider performance accurately on quality. Knowledge of these issues may help inform future measure alignment initiatives.

Prior studies of measure alignment identified differences in measure specifications in private and public payment and care delivery programs. An analysis by Mathematica Policy Research of eight cholesterol management and six heart failure quality measures used in government and privately sponsored public reporting programs found variation in each measure’s numerator and denominator characteristics.¹ Higgins et al.² found that 17 out of 301 Physician Quality Reporting System (PQRS) measures were identical or aligned with measures used by health plans. Findings from these studies suggest that lack of alignment occurs when public and private healthcare entities adopt similar (i.e., the same measure concept but different technical specifications) quality measures.

Studies have also indicated that the lack of alignment in measures may be attributed to differences in how public entities calculate and report performance rates for measures. For instance, an evaluation of state reporting and monitoring of healthcare-associated infections (HAIs) showed variations in the risk adjustment methods and in the way states report surgical site infection (SSI) and central line-associated bloodstream infection (CLABSI) measures.^{3,4} For SSI measures, states varied in the type and number of procedures reported. For CLABSI measures, states varied in their risk adjustment methodology and aggregation of performance rates. For both measures, states differed in the time lag between data collection and publication of results. Publicly reported measures that are not aligned greatly diminish the potential impact of measures by leading to instances of inadequate or inaccurate consumer comprehension of healthcare quality.⁶

Previous studies suggested that the lack of alignment in measures stems from various measurement barriers and differing needs of the entities implementing the measures. In a study by Damberg et al.,⁵ the authors conducted interviews of key informants from private and public organizations that use quality measures for accountability and quality improvement to explore patterns in their use of healthcare quality measures and describe factors that influence entities’ need to modify standard measure specifications. Interviewees expressed strong preference for

measures endorsed by the National Quality Forum (NQF) as well as measures aligned with other existing quality reporting initiatives. Nevertheless, they were also inclined to either modify existing measures or develop their own measures (i.e., “locally developed”) to overcome measurement barriers such as unavailable data sources or lack of measures focused on a specific topic of interest.

The NQF rigorously scrutinizes measures it endorses to ensure they are evidence-based, valid, reliable, feasible, and useful. Information regarding the reliability and validity of the locally developed measures is not available. In certain instances, the portfolio of NQF-endorsed measures may not address certain program needs, and these locally developed measures can be innovative and fill a specific need. Bazinsky and Bailit’s analysis of 48 measure sets⁹ used by 25 states in their various health programs showed that most state health programs modify measure specifications to target their population of interest, for ease of implementation, and to facilitate buy-in from providers. Comparisons of private and public sector measures to NQF-endorsed measures have been conducted and are outside the scope of this study.^{2,9} However, previous studies have not explored the alignment of measures used by various state programs, by the VHA and by CMS. This study aims to address this gap by examining alignment of measures used by public-sector entities, including CMS, states, and the VHA.

Using data collected for this study, the research team assessed similarities and differences between the measures used by CMS, states, and the VHA. For this study, alignment was defined as a measure focused on the same measure topic or condition as a CMS measure and that had the same measure data definitions for age range, target population, measurement period, data source, code set changes, rates reported, and other identifiable measure characteristics. The research team selected these measure data elements for evaluating alignment to parallel the analysis performed in previous measure alignment studies.^{2,9}

Objectives

There are two objectives that guided the analysis in this study:

1. Identify and describe the healthcare quality measures used by states and the VHA according to program type, measure condition or topic, and measure type.
2. Examine how well the quality measures used by states and the VHA are aligned (have the same measure specifications) with the measures used by CMS as of 2013.

Methods

Measures and Data Sources

This descriptive study compared the measures used by health departments and Medicaid agencies in 50 states and the District of Columbia, as well as measures used by the VHA with measures used in related CMS programs. See Table 4-1 for the list of CMS programs included, and Appendix i-4 for a hyperlink to a complete list of CMS measures included in this analysis. The list includes the NQF endorsement status, NQF number if endorsed, and both the measure title used by the CMS program and the measure title used by NQF.

Table 4-1: Programs Included in Analysis

Setting	Program	Abbreviation
 Hospital	Hospital Inpatient Quality Reporting Program	Hospital IQR Program
	Hospital Value-Based Purchasing Program	Hospital VBP Program
	Hospital Readmissions Reduction Program	HRRP
	Hospital-Acquired Condition Reduction Program	HAC Reduction Program
	Medicare and Medicaid Electronic Health Record (EHR) Incentive Program for Eligible Hospitals and Critical Access Hospitals	EHR EH
	Hospital Outpatient Quality Reporting Program	Hospital OQR Program
	Ambulatory Surgical Center Quality Reporting Program	ASCQR Program
	Inpatient Psychiatric Facility Quality Reporting Program	IPFQR Program
	Prospective Payment System-Exempt Cancer Hospitals Quality Reporting Program	PCHQR Program
 Ambulatory	Physician Quality Reporting System	PQRS
	Medicare Electronic Prescribing Incentive Program	eRx Incentive Program
	Physician Feedback Program	Physician Feedback Program
	Medicare and Medicaid Electronic Health Record (EHR) Incentive Program for Eligible Professionals	EHR EP
	Medicare Shared Savings Program	MSSP
	Physician Compare	Physician Compare
	Medicare Part C (Display or Star Ratings)	Part C
	Medicare Part D (Display or Star Ratings)	Part D
	Health Home Core Quality Measures ^{xxix}	Health Home
	State Demonstrations to Integrate Care for Dual Eligible Individuals Core Reporting Program ^{xxx}	Dual Eligible
	Core Set of Children’s Health Care Quality Measures for Medicaid and CHIP (Child Core Set)	Medicaid Child
Core Set of Health Care Quality Measures for Adults Enrolled in Medicaid (Medicaid Adult Core Set)	Medicaid Adult	
 Post-Acute	Nursing Home Quality Initiative	NHQI
	Home Health Quality Reporting Program	HH QRP
	End-Stage Renal Disease Quality Incentive Program	ESRD QIP
	Hospice Quality Reporting Program	HQRP
	Inpatient Rehabilitation Facilities Quality Reporting Program	IRFQR Program
	Long-Term Care Hospitals Quality Reporting Program	LTCHQR Program

^{xxix} The Home Health Core Quality Measures were included in this chapter for comparison with state and VHA measures.

^{xxx} The State Demonstrations to Integrate Care for Dual Eligible Individuals Core Reporting Program was included in this chapter for comparison with state and VHA measures.

The research team broadened the study conducted by Bailit Health Purchasing for the Buying Value Initiative, which examined alignment between state measures and NQF-endorsed measures.⁹ The research team obtained the data from Bailit Health Purchasing, collected in May 2013 and consisting of 1,367 measures, and revised Bailit’s measure alignment coding approach to conform to the research question. The Bailit study compared measures used in state programs with NQF-endorsed specifications or the standard specifications from a known measure developer.⁹ The research team further evaluated measures that used “standard” specifications (i.e., from a known measure developer) as to whether or not they were used in CMS programs.

A review of publicly available reports and documents from state government and VHA websites identified additional programs and measures. The research team created search syntax, shown below, to identify programs and measures. Search terms were combined with state-specific government site domain operators and entered into the Google search engine to identify relevant websites. The research team identified and contacted state and VHA staff via telephone or e-mail to solicit information regarding incomplete measure information found in state or VHA websites.

Search Terms

- ◆ “[state name/federal agency name*] AND quality health”
- ◆ “[state name/federal agency name*] AND quality measures”
- ◆ “[state name/federal agency name*] AND health quality measures”
- ◆ “[state name/federal agency name*] AND hospital measures”
- ◆ “[state name/federal agency name*] AND physician measures”
- ◆ “[state name/federal agency name*] AND nursing home measures”
- ◆ “[state name/federal agency name*] AND home health measures”
- ◆ “[state name/federal agency name*] AND Medicaid measures”
- ◆ “[state name/federal agency name*] AND accountable care organization measures”
- ◆ “[state name/federal agency name*] AND health homes measures”
- ◆ “[state name/federal agency name*] AND dual eligible measures”
- ◆ “[state name/federal agency name*] AND patient centered medical home measures”
- ◆ “[state name/federal agency name*] AND long-term care measures”
- ◆ “[state name/federal agency name*] AND dialysis measures”
- ◆ “[state name/federal agency name*] AND physician measures”
- ◆ “[state name/federal agency name*] AND foster children measures”
- ◆ “[state name/federal agency name*] AND CHIP measures”
- ◆ “[state name/federal agency name*] AND healthcare associated infection measures”
- ◆ “[state name/federal agency name*] AND developmental disabilities measures”
- ◆ “[state name/federal agency name*] AND correctional facilities health quality measures”

The data collection period for this study was November 1, 2013, through April 30, 2014. The data consisted of health quality measures that were being collected and/or used by the VHA,

state health department programs, and state Medicaid programs. Measures that assess structure, process, outcome, efficiency, cost/resource use, patient satisfaction, and access were included in the data. Out of the 7,005 state-used measures collected, 451 measures (6 percent) did not have adequate measure information to determine alignment and were excluded from the alignment analyses for Objective 2. Five hundred fourteen measures used by the VHA were included in the analysis; none were excluded.

Analysis

Objective 1—Describe the Healthcare Quality Measures Used by States and the VHA

The research team collected a total of 7,005 measures from states. In this analysis, the term “states” includes the 50 states and the District of Columbia. Measure sets obtained from Bailit were included in the data. The research team collected the following data elements for each measure:

- ◆ Measure name.
- ◆ Measure description.
- ◆ Measure specifications (as available, which include definitions for age range, target population, data source, code sets, and other identifiable measure characteristics).
- ◆ Measure steward.
- ◆ NQF number (if applicable).
- ◆ Data source or collection method.
- ◆ Measure use by program type.
- ◆ Measure purpose.
- ◆ Measure type.
- ◆ Measure topic/condition.

The research team categorized each state-used measure by state program (accountable care organization [ACO], patient-centered medical home [PCMH], Health Home, HAI reporting, Dual Eligible, Medicaid, Medicaid managed care organization [MCO], Medicaid behavioral health MCO [BH MCO], report card, and other).

The research team categorized each measure according to purpose (public reporting, quality improvement, payment, accreditation/licensing, and tiering).

The research team assigned each measure a condition or topic adapted from the taxonomy of measure attributes developed by the U.S. Department of Health and Human Services (HHS) Measure Inventory.¹⁰ The list of conditions or topics used to categorize the measures can be found in Appendix 4-1. “Condition” represents a broad medical condition or characteristic (e.g., Diabetes and Cardiovascular) and “topic” represents a non-medical condition or characteristic (e.g., Community Care Coordination/Transitions of Care and Screening) that the measure is assessing. The research team also grouped the measures by type as defined below:

- ◆ **Process:** A measure that focuses on a process that leads to a certain outcome, meaning that a scientific basis exists for believing that the process, when executed well, will increase the probability of achieving a desired outcome.
- ◆ **Outcome:** A measure that assesses the results that are experienced by patients who have received healthcare.
- ◆ **Intermediate Outcome:** A measure that aims to meet specific thresholds of health outcomes.
- ◆ **Structure:** A measure that assesses aspects of the healthcare infrastructure that are generally broad in scope and system-wide (for example, staffing level).
- ◆ **Efficiency:** A measure concerning the cost of care associated with a specified level of health outcome.
- ◆ **Patient Perspective:** A measure that focuses on a patient’s report concerning observations of and participation in healthcare.
- ◆ **Cost/Resource Use:** A measure of health services counts (in terms of units or dollars) applied to a population or event (broadly defined to include diagnoses, procedures, or encounters). A resource use measure is specified to count the frequency of defined health system resources; these measures may be further specified to apply a dollar amount (for example, allowable charges, paid amounts, or standardized prices) to each unit of resource use—that is, monetize the health service or resource use units.
- ◆ **Composite:** A measure that contains two or more individual measures, resulting in a single measure and a single score. Composite measures may be composed of one or more process measures and/or one or more outcome measures.

For each subsequent analysis, the research team counted each measure collected as an individual observation. For example, if a breast cancer screening measure was collected in 36 state Medicaid programs, it was counted 36 times. Throughout the analysis, the research team used this approach to account for variation in the way a measure was defined when used in various programs.

The research team collected 514 measures from the VHA website and VHA staff.^{xxx1} The research team collected the same data elements for each measure as collected for the state-used measures, as applicable. Separate analyses were conducted for the VHA measures.

Objective 2—Examine How Well the Quality Measures Used by the States and the VHA Are Aligned With Measures Used by CMS as of 2013

The research team compared state and VHA measures with measures used in CMS programs as of December 31, 2013. The CMS measures were from the CMS Measures Inventory,^{xxxii} a

^{xxx1} Only measures from national programs were collected from the VHA.

^{xxxii} The CMS Measures Inventory is a repository of measures used by CMS in its various reporting and payment programs. The inventory contains measure information for measures in a particular CMS program, including Measure Title, Description, Numerator, Denominator and Exclusion Statements; Measure Type; NQF-Endorsement Information; and Measure Steward. The Inventory is publicly available and can be accessed at <http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/CMS-Measures-Inventory.html>.

repository of measures used by CMS that is publicly available on the CMS website. *Federal Register* rules are the primary source used to populate the CMS Measures Inventory.

The research team evaluated alignment by reviewing the latest available measure information from state programs and the VHA and comparing this information to measures used by CMS as of December 31, 2013 (see Appendix i-4 for a hyperlink to a list of CMS measures used for comparison). The research team also used measures listed in the CMS core measure reporting requirements for state Dual Eligible demonstration projects.¹¹

The amount of information obtained through the environmental scan varied for each state program measure. If available, the research team used measure specifications (i.e., numerator, denominator, exclusion, and data collection method), which include definitions for measure characteristics such as age range, target population, data source, and code sets, to determine alignment. If measure specifications were not available, the research team used information contained in the measure description. The research team also examined state and VHA source documents to identify additional alignment information when specifications were not available. This included documentation from state programs and the VHA regarding measure data elements (e.g., measurement period, data source) or changes made to the specifications contained in a CMS Medicaid measure core set.

Classification of Measures by Alignment Types

The research team met with the Bailit researchers to gain an understanding of how Bailit applied the alignment criteria to the measures that they reviewed and applied the same approach to the additional measures collected for this study. As an important caveat, because data were included from the Bailit study, the research team did not check for inter-rater reliability. In addition, because no evidence-based procedures for assessing alignment were available, the procedure used to analyze the measures was refined as problems were identified.

Three reviewers categorized the measures that were collected. The research team gave each reviewer a detailed description and examples of each alignment type, program type, measure purpose, measure type, and measure topic/condition. A senior project staff reviewed the categorizations, and disagreements among reviewers were resolved after discussions with the project management team.

The research team used five alignment types, based on the Bailit⁹ study, to examine alignment measure data elements for each paired comparison of entities (i.e., CMS and states, and CMS and the VHA). The five alignment types used in the Bailit study and further refined by the research team in this study are:

1. ***Aligned***: A measure was categorized as aligned if the following criteria were met:
 - ◆ The measure focused on the same measure topic or condition as a CMS measure and had the same measure data definitions for each of the following elements: Age range, target population, measurement period, data source, code set changes, rates reported, and other identifiable data elements in the specification or measure description; OR
 - ◆ The measure had the same NQF ID number or measure developer as a CMS measure and had the same measure data definitions for each of the following elements: Age

- range, target population, measurement period, data source, code set changes, rates reported, and other identifiable data elements in the specification or measure description; OR
- ◆ The source document indicated that the measure was aligned with a measure used in a CMS program; OR
 - ◆ The source document indicated that the measure was based on an identified set of CMS measure specifications; OR
 - ◆ The source document indicated that the measure performance rate was based on the CMS reported rate.
2. **Similar:** A measure was categorized as similar if the following criteria were met:
 - ◆ The measure focused on the same measure topic or condition as a CMS measure and had different measure data definitions for one or more of the following elements: Age range, target population, measurement period, data source, code set changes, rates reported, and other identifiable data elements in the specification or measure description; OR
 - ◆ The measure had the same NQF ID or measure developer as a CMS measure and had different measure data definitions for one or more of the following elements: Age range, target population, measurement period, data source, code changes, rates reported, and other identifiable data element in the specification or measure description; OR
 - ◆ The source document indicated modifications to a measure in a CMS core measure set.
 3. **Locally Developed (for state measures only) or VHA-Developed:** A measure was categorized as locally developed if the measure was developed by the state and did not meet the criteria for “aligned” or “similar.” A measure developed by the VHA that did not meet the criteria for “aligned” or “similar” was categorized as VHA-developed. The locally developed and VHA-developed categories included composites or bundles of standard measures (i.e., from a known measure developer) created by states or the VHA and not used by CMS.
 4. **Other-Standard:** A measure was categorized as other-standard if the identified measure developer for the state- or VHA-used measure was an entity other than the state, VHA, or CMS, and the measure did not meet the criteria for “aligned” or “similar” and was not used by CMS.
 5. **Undetermined:** A measure was categorized as “undetermined” if the measure did not meet the criteria for “aligned,” “similar,” “locally developed or VHA-developed,” or “other-standard,” and the research team did not find documentation of the measure developer.

The research team assessed the overall alignment of state and CMS measures. To assess for overall alignment, state measures were compared with measures used in the programs listed in Table 4-1. VHA measures were compared to measures used in CMS hospital reporting, Nursing Home Quality Initiative (NHQI), and PQRS payment programs.

The research team selected the following state programs that are parallel to programs and/or measures implemented by CMS for alignment analysis: State Medicaid, HAI reporting, hospital report card, and nursing home pay-for-performance programs. The research team conducted state and CMS program measure comparisons as follows:

- ◆ State Medicaid measures were compared primarily to the Core Set of Children’s Health Care Quality Measures for Medicaid and CHIP (Medicaid Child) and Core Set of Health Care Quality Measures for Adults Enrolled in Medicaid (Medicaid Adult). Measures included in Dual Eligible, Health Homes, and ACO programs were compared with measures included in the CMS Dual Eligible Core Reporting Requirement, Health Homes Core Quality Measures, and the Medicare Shared Savings Program (MSSP). If the state measures were not included in the aforementioned CMS programs, a comparison using measures from PQRS and Medicare Part C (Display or Star Ratings when applicable) was performed. Accordingly, measures used in other CMS programs were used for comparison as appropriate.
- ◆ State HAI reporting programs were compared to HAI measures used in the following CMS programs: Hospital Inpatient Quality Reporting Program (Hospital IQR Program), End-Stage Renal Disease Quality Incentive Program (ESRD QIP), Long-Term Care Hospitals Quality Reporting Program (LTCHQR Program), Inpatient Rehabilitation Facilities Quality Reporting Program (IRFQR Program), and Prospective Payment System-Exempt Cancer Hospitals Quality Reporting Program (PCHQR Program). The research team did not compare the state measures with measures used in the Hospital-Acquired Condition Reduction Program (HAC Reduction Program) because identical hospital-acquired condition measures are used in the Hospital IQR Program.
- ◆ State hospital report card programs were compared to measures published on the CMS *Hospital Compare* website and measures implemented in the Hospital IQR Program and the Hospital OQR Program.
- ◆ State nursing home pay-for-performance programs were compared to the measures used in the Nursing Home Quality Initiative (NHQI) Program and with the measures posted on the CMS *Nursing Home Compare* website.

Examination of Similar Measures

Similar measures were examined to identify differences in the following measure specification characteristics:

- ◆ **Age:** Differences in the age inclusion criteria (e.g., all patients, child, adolescent, or adult).
- ◆ **Target population:** Differences in the target population (e.g., Medicare, Medicaid, Dual Eligible).
- ◆ **Measurement period:** Differences in the time period when data were collected for the measure, including start/end date or length of time.
- ◆ **Rate reported:** Differences in how a rate was reported (e.g., single rate, age-stratified rates).

- ◆ **Data source:** Differences in the source of data for abstracting a measure (e.g., paper medical records, claims, electronic health records).
- ◆ **Other measure data element differences:** Other differences in measure data elements not described above (e.g., exclusion criteria, risk adjustment, number of visits).

This report presents descriptive information on state program and VHA measures, including numbers and percentages for program/measure background characteristics. Created using ArcGIS software, the choropleth map in Appendix 4-3 illustrates the geographical variation in the number of measures collected from each state. The legend values for classification breaks are based on Jenks Natural Breaks method.¹² This approach uses an iterative algorithm to assign data optimally to classes such that the variances within classes are minimized, while the variances between classes are maximized.

Limitations

This study recognized that the populations served by CMS, states, and the VHA are different. Therefore, total measure alignment was not expected.

This study did not explore whether the data collection dates for the measure rates published in state and CMS report card programs were aligned. This aspect of alignment can be explored in future studies.

The analyses were limited by the scope of measures included in the data. The research team only evaluated measures used by states and the VHA for alignment with measures used by CMS. This study did not examine the alignment of measure specifications across CMS programs, and it also did not explore alignment in measures implemented in CMS and private healthcare delivery programs. However, findings from other studies have provided evidence regarding alignment between CMS and private programs. For example, there was variability in specific measures that health plans used to evaluate quality related to upper respiratory illness, cost of care, pulmonary conditions, and asthma when compared with measures used in MSSP.²

This study was also limited by the search strategy used to identify state-used measures. The environmental scan focused primarily on state Medicaid agency and health department websites. As a result, measure lists not identified using this approach may exist.

Although the research team used other means such as e-mail and telephone communication to obtain measures from states, response rates to these types of solicitations were low. The research team received 18 responses out of 78 inquiries (23 percent) for measure information from the states. As a result, the research team obtained full measure specifications for less than half of the measures in this study. In those cases, the research team used wording from the measure description to determine alignment. Although the research team obtained additional details regarding state-used measures from the source documents, other measure characteristic differences unaccounted for in the current analysis may exist, such as data collection dates used to calculate measure performance rates and data collection mechanisms used to report measure performance rates according to demographic groups (e.g., race, ethnicity, and primary language).

State health officials did not validate the list of measures collected. Finally, national and local efforts intended to bring measures into closer alignment are in progress; at the time these study results become available, measures may have changed.

Results

Objective 1—Identify and Describe the Healthcare Quality Measures Used by the States and the VHA

The research team found 7,519 state and VHA measures. Table 4-2 shows an overview of the total number of state and VHA measures.

Table 4-2: Overview of the Number of Measures for States and the VHA, and Number of State Programs

Entities	Total	Range
State Measures	7,005	16–391
VHA Measures	514	–
Total Measures	7,519	–

State Measures

The environmental scan identified 271 healthcare programs across states that are using quality measures. State-used measures totaling 7,005 were collected from the state programs, of which 1,775 are unique or unduplicated. The total number of measures varied across programs and states. The number of measures per state ranged from 16 to 391 measures. Appendix 4-2 shows the distribution of the number of programs found per state. Appendix 4-3 presents the geographical variation in the number of measures found across the states. Of the 7,005 state-used measures identified, 451 measures (6 percent) did not have adequate measure information to determine alignment. A total of 514 measures were identified from VHA programs.

Types of State Programs

All 50 states and the District of Columbia use quality measures across different programs. The number of programs found per state ranged from two to 13 programs. The number of measures collected also varied according to the type of program using the measures. The measures used in Medicaid programs (i.e., Medicaid FFS, MCO, BH MCO, Dual Eligible, PCMH, Health Home, and ACO programs) accounted for over half of the measures (57 percent) identified in state programs. State report card program measures accounted for 1,438 (21 percent) of the state-used measures. Appendix 4-4 shows the total number of measures according to the type of state program.

Purposes of Measure Use by State Programs

The research team identified 401 discrete purposes across the 271 programs included, with 119 programs indicating two or more purposes for measure use. Among the programs, quality improvement (n=162) and public reporting (n=103) were the purposes most frequently indicated. Appendix 4-5 shows an overview of the number of programs according to the purpose type.

Topics/ Conditions Addressed by State Measures

The research team found that in programs across the states, these 10 conditions or topics were measured most frequently: Patient safety (11 percent), cardiovascular (10 percent), mental health and substance-related care (6 percent), respiratory (6 percent), diabetes (6 percent), preventive care (6 percent), patient experience (5 percent), utilization (5 percent), surgical procedures (5 percent), and general health services administration (4 percent). Measures focused on these 10 conditions or topics comprised 63 percent (n=4,394) of the total number of state-used measures in the data. Appendix 4-6 shows the distribution of conditions or topics that were the focus of the state-used measures.

Measure Types Used by States

State-used measures varied in the aspect of quality they assessed. The research team identified eight different types of measures across all state-used measures included in the analysis (Appendix 4-7). Most measures fell under the process measure category, which accounted for 66 percent of the overall state-used measures in the data. Twenty percent of the measures evaluated outcomes of care. Measures focused on patient perspective, intermediate outcomes, structure, composite, efficiency, and cost/resource use accounted for 14 percent of the state-used measures.

VHA Measures

The research team collected 514 measures from the VHA. The VHA measures addressed these 10 conditions or topics most frequently: Cardiovascular (15 percent), health services administration—quality improvement (15 percent), patient safety (12 percent), mental health and substance-related care (10 percent), health services administration—patient experience (9 percent), diabetes (6 percent), surgical procedures (6 percent), general health services administration (4 percent), and mortality (4 percent), and respiratory (4 percent). Appendix 4-8 shows the distribution of conditions or topics that were the focus of the VHA measures.

The research team found eight different types of measures across the VHA measures (Appendix 4-9). Forty percent of the VHA measures were process measures, followed by structural (21 percent), outcome (13 percent), and intermediate outcome measures (9 percent). Measures focused on cost/resource use, efficiency, patient perspective, and composite accounted for 17 percent of the overall VHA measures.

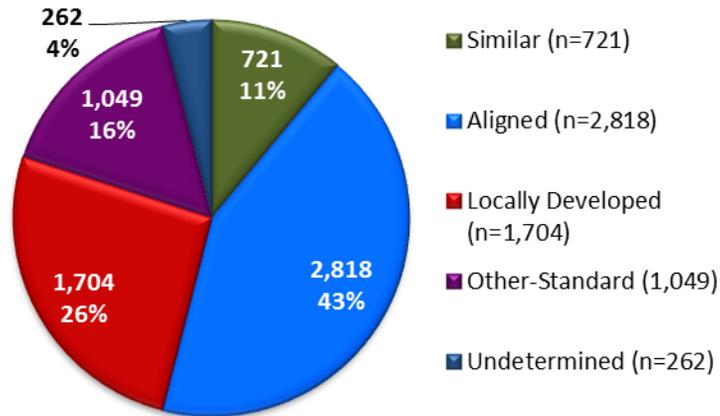
Objective 2—Examine How Well the Quality Measures Used by the States and the VHA Are Aligned With Measures Used by CMS as of 2013

Overall—State Programs Measures Alignment With CMS Measures

The research team examined alignment of state-used measures and CMS-used measures. Of the 7,005 state-used measures collected, 6,554 measures (94 percent) were included in this analysis; 451 measures (6 percent) did not have adequate measure information to determine alignment and were excluded from the alignment analysis. Forty-three percent of 6,554 state-used measures aligned with measures used by CMS in one or more of its quality reporting and payment

programs; 11 percent measured similar concepts. However, 26 percent of state-used measures were locally developed and not aligned or similar; an additional 16 percent were not aligned or similar (other-standard) measures that were developed by parties other than CMS and the states and not used in any CMS program. Figure 4-1 shows a summary of state-used measures by alignment type.

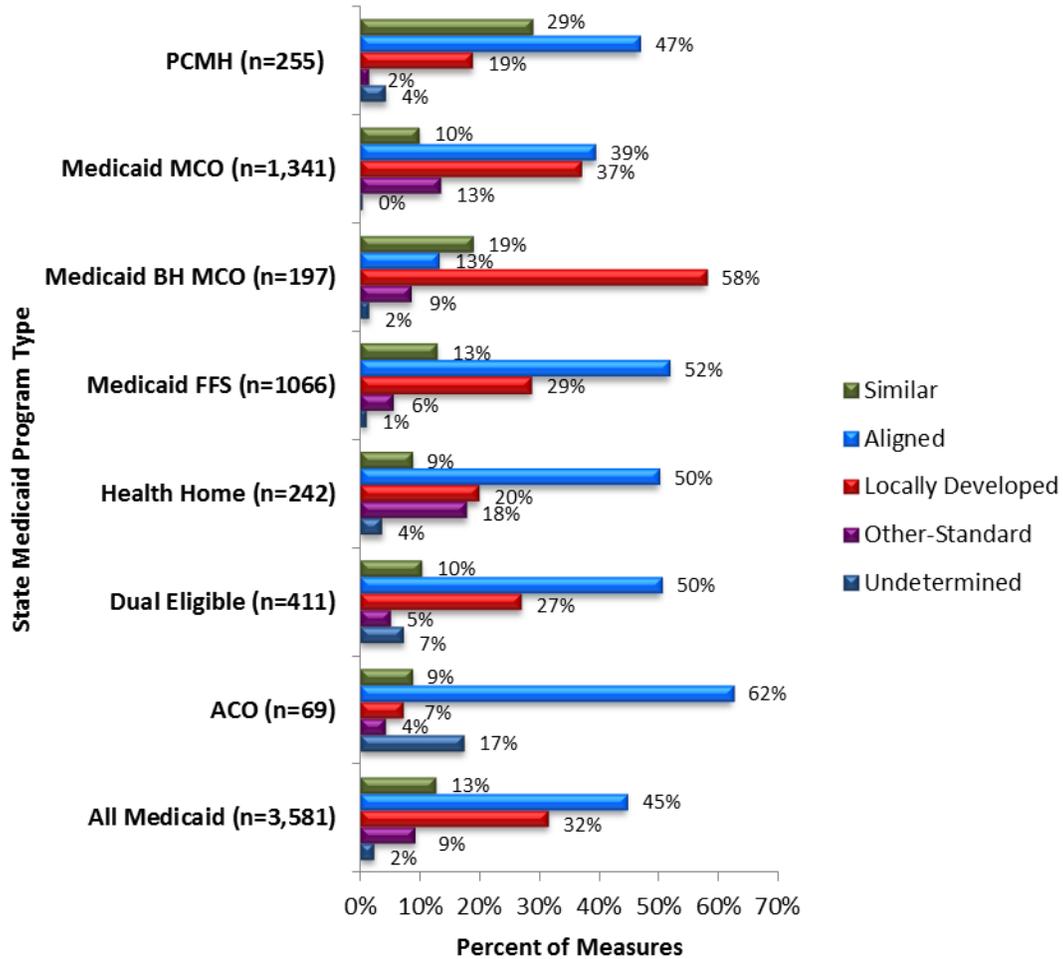
Figure 4-1: Summary of State-Used Measures by Alignment Type (n=6,554)



State Medicaid Program Measures Alignment With CMS Measures

Figure 4-2 includes a summary of alignment of state Medicaid measures and CMS measures and shows variability in the alignment of measures used by CMS and state Medicaid programs. Analysis showed that 45 percent of 3,581 state Medicaid measures aligned with measures used in CMS programs, while 13 percent were similar. Locally developed measures that were not aligned or similar were present in 32 percent of the state Medicaid measures. These measures focused on conditions such as prenatal care and attention deficit hyperactivity disorder (ADHD), as well as general health services administration measures. Other non-aligned or non-similar measures developed by parties other than CMS or states and not used in any CMS program (i.e., other-standard) were present in 9 percent of the state Medicaid measures. Examples of these measures included measures that assess lead screening in children and board certification measures, developed by the National Committee for Quality Assurance (NCQA).

Figure 4-2: Percentage of Measures Found in State Medicaid Programs by Alignment Type



In examining alignment within specific Medicaid programs, the research team found that Medicaid BH MCO programs included the least number of aligned measures. The research team found higher percentages of aligned measures in ACO (62 percent), Dual Eligible (50 percent), Health Home (50 percent), and Medicaid FFS (52 percent) programs compared to measures in PCMH (47 percent), Medicaid MCO (39 percent), and Medicaid BH MCO (13 percent) programs. In general, in programs for which CMS identified a measure core set, such as Medicaid FFS, Dual Eligible, and Health Home programs the percentages of aligned measures were higher compared to measures that were similar.

Further analysis of measures used in state ACO programs showed that of the overall number of aligned measures (n=43), 25 percent (n=11) were aligned with measures used in the MSSP. However, the remaining measures aligned with measures used in other CMS programs, such as PQRS and Part C.

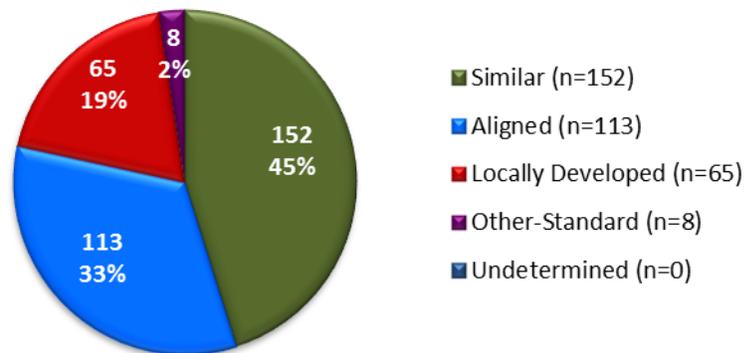
Examination of Similar Measures Used in State Medicaid Programs

The research team identified 495 distinct measure characteristic differences across similar measures used in the state Medicaid programs. There were 446 state Medicaid measures that were similar to CMS measures. Of these, 10 percent were categorized as similar due to differences in two or more measure characteristics; 24 percent varied in the way the rates were reported, and 32 percent exhibited differences in measure data elements. For example, in the *Chronic Obstructive Pulmonary Disease Admissions*^{xxxiii} measure, the research team noted differences in the way state Medicaid and CMS programs reported performance rates. For this measure, a state Medicaid FFS program elected to report an aggregate rate rather than the age-stratified rate indicated on the Medicaid Adult core set. In other instances, states may have used age stratifications for reporting other measures that differed from the Medicaid Adult recommended age stratifications. State Medicaid measures were also found to contain exclusions and other criteria not found in similar CMS measures. For example, one state excluded dual eligible individuals from its Medicaid measures. In a state PCMH program using the *Colorectal Cancer Screening* measure (NQF #0034), a patient must have two or more visits to his or her provider to be included in the denominator. This measure was also included in the PQRS Program, but the PQRS measure specifications require a patient to have one or more visits with his or her provider to be included in the measure denominator.

State HAI Reporting Measure Alignment With CMS Measures

The research team found 338 measures used across HAI reporting programs in 35 states. States aligned one-third (33 percent) of their HAI program measures with HAI measures developed by the Centers for Disease Control and Prevention (CDC) and implemented by CMS in the Hospital IQR Program and CMS *Hospital Compare*. However, 45 percent of the HAI measures used by states were similar to measures used in these CMS programs. In addition, 19 percent of the state-used HAI measures were locally developed, and 2 percent were other-standard measures not used in CMS programs. Figure 4-3 shows the percentage of state HAI measures according to alignment type.

Figure 4-3: Percentage of Measures Found in State HAI Programs by Alignment Type (n=338)



^{xxxiii} NQF-endorsed title *PQI 05: Chronic Obstructive Pulmonary Disease (COPD) Admission Rate* (NQF #0275).

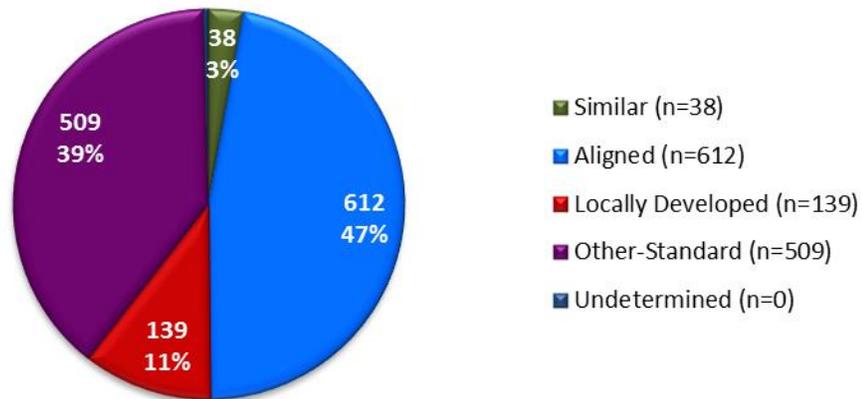
Examination of Similar Measures Used in State HAI Reporting Programs

Of the 153 state HAI measures that were similar to the CMS HAI measures, 93 percent contained differences that pertain to reported rates. For example, many states reported other types of surgical procedures for SSI measures, such as coronary artery bypass graft, hip and/or knee prosthesis surgeries, and laminectomy. CMS collects SSI rates for colon surgery and abdominal hysterectomy for the Hospital IQR, Hospital VBP, HAC Reduction, and PCHQR programs. Similarly, the research team observed differences in the way states and CMS reported CLABSI measures. Most states reported separate CLABSI rates for different intensive care unit (ICU) types, such as adult ICU, neonatal ICU, neurosurgical ICU, and trauma ICU. In comparison, CMS has implemented the measure in the Hospital IQR, Hospital VBP, HAC Reduction, LTCHQR, and PCHQR programs that reports one rate and does not specify type of ICU. Finally, a large proportion of state-used HAI measures were categorized as locally developed. Many of these measures were non-standardized infection ratio (SIR) for CLABSI, SSI, methicillin-resistant *Staphylococcus aureus* (MRSA), catheter-associated urinary tract infection (CAUTI), *Clostridium difficile* (*C. diff*), and *Escherichia coli* (*E. coli*).

State Hospital Report Card Measure Alignment With CMS Measures

The research team found hospital reporting programs in 21 states. Less than half (47 percent) of the 1,303 measures used across the state hospital report card programs were aligned with CMS measure, and 39 percent were other-standard measures (i.e., measures from a known developer that were not implemented in CMS hospital programs). Other-standard measures included mortality, patient safety, and volume indicators for hospital procedures developed by the Agency for Healthcare Research and Quality (AHRQ). Analysis also showed that two heart attack mortality measures (one developed by CMS and the other developed by AHRQ) appeared in the same hospital report cards in seven states. Figure 4-4 presents the percentage of state hospital report card measures according to alignment type.

Figure 4-4: Percentage of Measures Found in State Hospital Report Cards by Alignment Type (n=1,303)



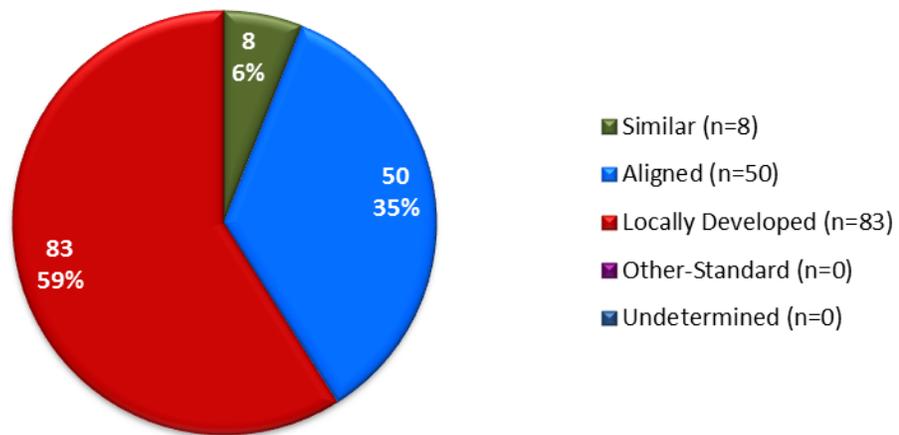
Examination of Similar Measures Used in State Hospital Report Card Programs

Three percent (n=38) of the state hospital report card measures were similar to measures used in the CMS hospital programs. A majority of the differences between state measures and CMS measures were due to differences in the target population and rates reported. Forty-seven percent (n=18) of the similar state hospital report card measures contained differences that pertain to other changes in target population. For example, many states use the standard specifications for AHRQ Quality Indicators, which target a geographical population. In contrast, many CMS specifications include only the Medicare population. Thirty-nine percent (n=15) of the measures in the hospital report card programs differed in which rates were reported. For example, the state SSI measures included rates for surgical procedures other than colon surgery and abdominal hysterectomy. CMS uses *American College of Surgeons – Centers for Disease Control and Prevention (ACS-CDC) Harmonized Procedure Specific Surgical Site Infection (SSI) Outcome Measure (NQF #0753)* that is applied to two operative procedures, colon surgeries and abdominal hysterectomies, only in the Hospital IQR, Hospital VBP, HAC Reduction, and PCHQR programs.

State Nursing Home Pay-for-Performance Measure Alignment With CMS Measures

The research team found nursing home pay-for-performance programs in nine states. Thirty-five percent of the 141 measures used across the state nursing home pay-for-performance programs aligned with measures implemented in CMS nursing home programs. Developed by CMS, these measures are based on the Minimum Data Set. Analysis of state nursing home pay-for-performance measures also showed that 59 percent of all measures in this set were locally developed (Figure 4-5). Six of the nine states in this analysis had one or more locally developed measures included in their programs, and seven out of nine states had one or more measures aligned with CMS measures. State nursing home locally developed measures focused on areas such as staff retention or turnover, staff nursing education, and elements of culture change that pertain to dining and bathing.

Figure 4-5: Percentage of Measures Found in State Nursing Home Pay-for-Performance Programs by Alignment Type (n=141)



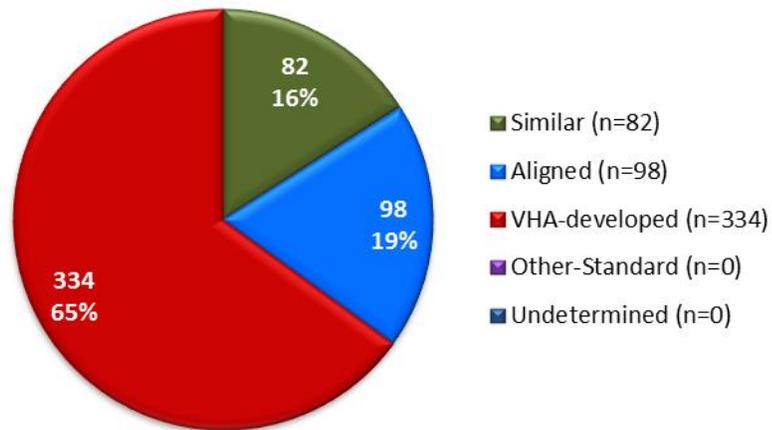
Examination of Similar Measures Used in State Nursing Home Pay-for-Performance Programs

A small percentage of state nursing home measures were modified (i.e., similar) CMS measures—this was observed in two states. State modifications to the CMS measures include changes in other measure data elements, such as look-back periods and exclusions. One example involves the measure used in a state nursing home pay-for-performance program assessing pain in short-stay residents. The state-used short-stay pain measure had a seven-day look-back period compared to the CMS measure *Percent of Residents who Self-Report Moderate to Severe Pain (Short Stay)*(NQF #0676) five-day look-back period used in the CMS Nursing Home Quality Initiative. Similarly, a state using a measure that evaluated the presence of a catheter in long-stay residents used an exclusion criterion pertaining to residents in hospice care and residents with an end-stage prognosis, which are not included in the related CMS measure *Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay)* (NQF #0686) used in the CMS Nursing Home Quality Initiative.

Veterans Health Administration Measure Alignment With CMS Measures

Figure 4-6 shows the percentage of VHA measures by alignment type. Nineteen percent of the 514 VHA measures aligned with CMS measures, while 16 percent were similar. Nearly two-thirds (65 percent) of the VHA measures in this analysis were not aligned or similar to measures used by CMS and were VHA-developed. The VHA locally developed measures are focused on a variety of topics, including health services administration (e.g., employee satisfaction and patient safety culture); communicable disease (e.g., tuberculosis); and mental health (e.g., post-traumatic stress disorder and suicide). The VHA-developed measures also included composite measures focused on topics such as behavioral health, diabetes, tobacco cessation, ischemic heart disease, and cancer prevention. The composite measures reported the absolute difference for the measures between males and females, and between whites and non-whites.

Figure 4-6: Percentage of VHA Measures by Alignment Type (n=514)



Examination of Similar Measures Used in Veterans Health Administration Programs

Ninety-nine discrete characteristic differences in VHA measures were found when compared with CMS measures. Of the 82 VHA measures that were similar to CMS measures, 17 have differences in two or more measure characteristics. The most frequently noted differences between VHA and CMS measures pertained to the target population and exclusion criteria. For example, the VHA used a Healthcare Effectiveness Data and Information Set (HEDIS^{®xxxiv})-based measure, *Diabetes: Eye Exam^{xxxv}*, which is similar to the PQRS and EHR EP measure. In addition to using the measure for the entire population with patients, the VHA uses this measure for a limited population of patients with diabetes that have spinal cord injury and disorders. The HEDIS-based VHA eye exam measure also used exclusion criteria, such as terminal illness or enrollment in a hospice program, which were not included in the CMS exclusion list for this measure.

Discussion

This study examined alignment between state- and CMS-used measures, as well as between VHA and CMS measures. The results indicated that state program measures had a higher percentage (43 percent) of aligned measures than the VHA (19 percent). The States and the VHA use measures that are similar to ones in CMS programs, but have differences in the specifications. While both state programs and the VHA programs include a large number of locally developed measures, the VHA had a higher percentage, 65 percent, compared to 26 percent of the measures used in state programs. Understanding the underlying reasons for these differences may facilitate efforts to increase the alignment across these public sector programs.

Analysis of state measures showed that alignment varied according to program type. Across measures used in state Medicaid programs, the overall percentage of aligned measures (45 percent) was higher compared to similar measures (13 percent). Similar distribution of alignment types was also found in specific Medicaid programs, especially in programs where CMS has measure core sets, such as Medicaid FFS, Dual Eligible, and Health Home programs. The analysis showed that differences exist in the measure specifications used in these programs as compared to the Medicaid core set specifications. This finding was anticipated because core sets, such as those developed by the Subcommittee on Children’s Healthcare Quality Measures for Medicaid and CHIP Programs (SNAC) for the Medicaid Child, did not recommend that measures in the core set be implemented according to core set specifications. As an alternative, denominator modifications are allowed so that measures are feasible for use by various types of Medicaid programs.¹³ As a result, study findings indicated that the way state Medicaid programs report measure rates varies. Findings also showed that state Medicaid measures contain differences in the target population and other measure data elements when compared with Medicaid core set specifications.

^{xxxiv} HEDIS[®] is a registered trademark of the National Committee for Quality Assurance (NCQA).

^{xxxv} Similar to *Comprehensive Diabetes Care: Eye Exam (retinal) performed* (NQF #0055).

The research found that nearly two-thirds (32 percent) of the measures in various state Medicaid programs were locally developed. The highest percentage of locally developed measures was found in state Medicaid behavioral health programs (58 percent), followed by Medicaid MCO (37 percent), Medicaid FFS (29 percent), Dual Eligible (27 percent), and PCMH (19 percent). State Medicaid locally developed measures focused on mental healthcare and substance-related conditions, such as ADHD and perinatal depression, as well as health services administration, which includes measures that evaluate whether a patient enrolls in a type of service.

In state HAI reporting programs, the percentage of similar measures (45 percent) was higher when compared to aligned measures (33 percent). State- and CMS-used HAI measures differed in the type of surgical procedure and/or hospital units included in their published measure performance rates. For example, 15 out of 35 states with HAI reporting programs reported on colon surgery, which is an important procedure to measure nationally due to high SSI rates. The low number of states reporting SSI related to colon surgeries echoes findings from previous literature that identified significant variation in the reporting practices for SSI and CLABSI rates in state HAI programs.^{3,4} However, the authors from these studies were not able to determine what prompts certain states to legislate reporting of procedures for SSI and care setting locations to report for CLABSI.

State hospital report card programs had a higher percentage of aligned measures compared to similar measures. Forty-seven percent of the state-used hospital report card measures aligned with measures used in CMS hospital quality initiatives, while 3 percent were similar. The integration of CMS *Hospital Compare* measures in state hospital report cards ensures consistency of information in state and CMS public reporting systems for consumers.

Thirty-nine percent of state hospital report card measures are other-standard measures not used by CMS Hospital IQR. These other-standard measures consist of AHRQ-developed hospital volume, mortality, and patient safety indicators that states use to complement the CMS *Hospital Compare* information. For example, seven state hospital report cards concurrently reported *Hospital 30-day, All-Cause, Risk-Standardized Mortality Rate (RSMR) Following Acute Myocardial Infarction (AMI) Hospitalization for Patients 18 and Older* (NQF #0230, developed by CMS) and *Acute Myocardial Infarction (AMI) Mortality Rate* (NQF #0730, developed by AHRQ). These two mortality measures used different approaches to calculate mortality. NQF #0230 includes all deaths within 30 days of admission for patients with AMI, while NQF #0730 includes only mortality during the course of hospitalization. In this instance, mortality information in certain state report cards contained redundant information. These findings are similar to those from previous studies that examined public reporting of hospital quality that identify inconsistencies in mortality measure approaches (i.e., CMS versus AHRQ) selected by states in their public reporting programs.¹⁴

The largest percentage of measures used in state nursing home pay-for-performance programs were locally developed (59 percent) and addressed topics such as staff retention or turnover, staff nursing education, and elements of culture change that pertain to dining and bathing. However, no CMS nursing home measures addressed these topics which are important to the states. While the state programs are limited to Medicaid patients, these topics may be applicable to the Medicaid and Medicare populations included in the CMS measures.

Nineteen percent of measures used by the VHA aligned with CMS-used measures. These measures consisted of inpatient hospital measures, which included patient experience surveys used by CMS and reported on the CMS *Hospital Compare* website. Fifteen percent of VHA-used measures were similar to CMS-used measures and were primarily physician measures that included different data abstraction methods, exclusion criteria, and the target population from CMS-used measures. The percentage of VHA-developed measures was high, which may reflect the VHA being a contained healthcare system with EHR capabilities that allow for the collection of data and customization of measures that address combat- or service-related healthcare needs.

Results from other studies^{1, 5} suggested other factors that may help account for the lack of alignment in measures used by different entities. These include perceptions held by groups and organizations that gaps exist in the coverage of existing measures as well as the desire to fine-tune measurement specifications to match the needs of an entity's particular population.

Findings show that the states and the VHA develop their own measures, which were widely used by these entities (26 percent and 65 percent, respectively) in various programs. These findings may be due to modest differences in the population as well as the healthcare needs of the population served by these entities. Different policy decisions made by states, the VHA, and CMS may contribute to differences in the measures. For example, CMS has also modified the target population of a measure to focus on Medicare beneficiaries and has made changes to other measure data elements to meet its reporting requirement needs. The underlying factors contributing to the differences in similar measures warrant further examination.

Conclusions

Key Findings

Less than half of the quality measures used by states and VHA are aligned (i.e., measures used by other entities that use specifications consistent with specifications used by CMS) with CMS measures. State Medicaid and state hospital report cards are more closely aligned with CMS measures than measures developed by the VHA.

This study found that states and the VHA implemented quality measures focused on various conditions such as those related to patient safety, cardiovascular, and mental health. However, measures used by states, the VHA, and CMS differed in target population and other specific measure data elements, such as exclusion criteria and how the measure rates were reported.

Analyses of the measures used by state Medicaid behavioral health MCO and nursing home pay-for-performance programs and the VHA showed that over half of the measures are developed by the states and the VHA and are not used in CMS programs. High percentages of such unaligned and non-similar measures in state and VHA programs may indicate a lack of available measures to address priority measurement areas for these entities.

Actions to Consider

This study found variations in specifications or measure data elements, reporting requirements, sampling methodologies, and reporting deadlines. Additional research is needed to determine to what extent further alignment would benefit providers and patients.

Action to Consider: Investigate the underlying reasons for measure specification variations between CMS and other entities (states, other federal agencies, private sector, etc.). Understanding these underlying factors for the measure specification variations will provide insights for further refinement of measure alignment principles and strategies.

Given the differences in target population, data sources, and program needs, the principle that measure specifications should be entirely uniform may not always be the best policy in all situations. Therefore, the principle of measure alignment may need to be redefined in a way that prioritizes the measure data elements that are most important for achieving comparability. Addressing these issues is key to moving the measure alignment agenda across public and private entities.

Action to Consider: Determine what degree of alignment with state and federal programs would benefit patients and providers. Measure alignment not only reduces provider burden but also supports a multi-payer approach to transforming healthcare. However, the principle that measure specifications should always be uniform should be re-examined and potentially redefined in a way that prioritizes the elements that are most important for comparability.

An in-depth review of non-CMS-developed measures used by the states and the VHA identified in this study may suggest innovative approaches to measuring and reporting quality information. The states and the VHA have deployed non-CMS measures to meet the specific needs of a particular program or region that are innovative and address topics that CMS and other entities can adopt. The states and the VHA serve as testing grounds for such measures. CMS can assess these non-CMS measures used by the states and the VHA to determine their applicability for use in CMS programs.

Action to Consider: Review key sources of non-CMS measures used by states and the VHA as a component of environmental scans conducted for new measure development efforts. An in-depth review of these measures may suggest innovative approaches to measuring and reporting quality information that could be applied broadly and adopted within CMS programs. This may be particularly relevant for electronic clinical quality measures developed for the VHA.

Reference List

- (1) O'Neil S, Schurrer J, Simon S. *Environmental Scan of Public Reporting Programs and Analysis*. Final Report. Cambridge, MA. 2010; Final Report. Mathematica Reference Number: 06738.500.
- (2) Higgins A, Veselovskiy G, McKown L. Provider performance measures in private and public programs: achieving meaningful alignment with flexibility to innovate. *Health Aff (Millwood)*. 2013;32(8):1453-1461.
- (3) Aswani MS, Reagan J, Jin L, Pronovost PJ, Goeschel C. Variation in public reporting of central line-associated bloodstream infections by state. *Am J Medical Qual*. 2011;26(5):387-395.
- (4) Makary MA, Aswani MS, Ibrahim AM, Reagan J, Wick EC, Pronovost PJ. Variation in surgical site infection monitoring and reporting by state. *J Healthc Qual*. 2013;35(2):41-46.
- (5) Damberg C, Sorbero M, Lovejoy S, et al. *An Evaluation of the Use of Performance Measures in Health Care*. Santa Monica, CA: National Quality Forum; 2011.
- (6) National Quality Forum. *Measure Applications Partnership Strategic Plan: 2012-2015*. Washington, DC: October 2012.
- (7) Sood N, Higgins A. Posing a framework to guide government's role in payment and delivery system reform. *Health Aff (Millwood)*. 2012;31(9):2043-2050.
- (8) Centers for Medicare & Medicaid Services. *CMS Strategy: The Road Forward, 2013-2017*. Baltimore, MD: Centers for Medicare & Medicaid Services; March 2013.
- (9) Bazinsky K, Bailit M. *The Significant Lack of Alignment Across State and Regional Health Measure Sets*. 2013.
- (10) Agency for Healthcare Research and Quality, Department of Health and Human Services. HHS Measures Inventory - Template of Measure Attributes. Agency for Healthcare Research and Quality 2014 January 1; Available at: URL: <http://www.qualitymeasures.ahrq.gov/hhs/about/template.aspx>.
- (11) Centers for Medicare & Medicaid Services. *Medicare-Medicaid Capitated Financial Alignment Model Reporting Requirements*. 2014.
- (12) Jenks GF, Caspall FC. Error on choroplethic maps: definition, measurement, reduction. *Annals of the Association of American Geographers*. 1971;61(2):217-244.
- (13) Initial Core Set of Children's Healthcare Quality Measures: Summary Report. January 2011. Agency for Healthcare Research and Quality, Rockville, MD. <http://www.ahrq.gov/policymakers/chipra/listtable.html>
- (14) Ross JS, Sheth S, Krumholz HM. State-sponsored public reporting of hospital quality: results are hard to find and lack uniformity. *Health Aff (Millwood)*. 2010;29(12):2317-2322.

Chapter 5—CMS Measures: Populations Reached



Questions on Reach

1. What populations are reached by the quality measures used in Medicare programs?
2. How frequently are specified populations excluded from quality measures?

Abstract

Background: To be an effective lever in improving healthcare, Centers for Medicare & Medicaid (CMS) quality measures must reach the appropriate beneficiaries. The reach of a quality measure is evaluated by identifying populations that are included or intended to be included in the target population of a measure. This study examines Medicare quality measure inclusion and exclusion criteria and identifies opportunities for improved measure reach.

Methods: This descriptive study examines specifications for 754 Medicare quality measures in use or finalized in a rule for future use as of December 31, 2013. The research team analyzed specifications by medical condition or topic (e.g., mortality or readmission), age criteria, and the type of population by payer for each measure. Measure denominator exclusion criteria were categorized to identify populations excluded from the measure, and frequency analyses were conducted to determine the extent to which populations were either included or excluded in Medicare program measures. The Kappa coefficient statistical approach was used to determine inter-rater reliability for categorizing exclusions. The research team achieved substantial agreement (a Kappa coefficient of 0.7 or greater) for categorizing measure exclusions.

Results: Findings showed that 33 conditions and 77 sub-conditions were represented in Medicare measures. Medicare program measures addressed 18 of the 20 high-impact Medicare conditions. Age category evaluations showed that 94 percent of the measures reached individuals below age 65. Payer type evaluation demonstrated that 88.5 percent of Medicare measures included Medicare fee-for-service (FFS) patients, 49.6 percent of the measures included Medicaid patients, 42.3 percent of the measures included Medicare Part C enrollees, and 31.4 percent of the measures

included patients with other payer sources. The most commonly found exclusions included medical diagnosis or other clinical reasons (29.4 percent), end-of-life care (24.3 percent), provider discretion (25.6 percent), and patient refusal (19.1 percent). Within programs, provider discretion and patient transfer to or from another facility were the most commonly found exclusions. Provider discretion was an exclusion criterion in 32.6 percent of Hospital Inpatient Quality Reporting Program (Hospital IQR Program) measures and 34.6 percent of physician practice measures. Patient transfer was an exclusion criterion in 81.5 percent of Hospital IQR Program measures and 38.9 percent of Prospective Payment System-Exempt Cancer Hospitals Quality Reporting Program (PCHQR Program) measures. Exclusions related to high-risk conditions (totally dependent and comatose patients) and psychiatric conditions comprised 5.3 percent and 3.7 percent of the measures reviewed, respectively. Exclusions based on medical diagnosis were found in 41.7 percent of outcome measures and 25.8 percent of process measures, while exclusions based on provider discretion were found in 38 percent of process measures and 3 percent of outcome measures.

Conclusions: Medicare program measures reach a variety of populations with different medical conditions and healthcare coverage. Although the research team found a number of exclusions in the measures, no populations were consistently excluded across programs. However, intra-program exclusions (e.g., provider discretion, transfers) may warrant further investigation. Future efforts to broaden the reach of quality measures may include further development of measures that specifically target excluded populations or generally minimize exclusions.

Background

This study examines the first of five elements in the RE-AIM framework, which scholars call “reach.”¹ This element assesses the extent to which individual populations are represented within the measured population(s) and reached with particular quality programs and initiatives. The reach of a quality measure is evaluated by identifying populations that are included or intended to be included in the target population of a measure. Identifying populations excluded from a measure provides additional information about the reach of a measure.

Elements of a Measure

Each measure is specified with a set of data elements required to collect and calculate rates for the measure, including but not limited to numerator, denominator, target population, excluded population (exclusions), data source and data collection strategy, sampling, and measure algorithm. Measure developers strive to ensure that measure specifications address the salient aspects of care and are optimally quantified to include those patients intended to be included, while excluding those intended to be excluded. To achieve this, it is imperative that the measure development process includes scientific rigor, which entails testing measure specifications for reliability and validity.

Reach of a Measure—Target Population

Having a well-defined target population is a critical approach to ensure validity in the measures. The target population of a measure refers to the denominator or population sample meeting specified criteria of the measure.² In defining target populations, measure developers assign inclusion and exclusion criteria that are supported by evidence.³ For measures such as outcome measures, it may be necessary to apply risk adjustment methods to statistically offset (or adjust) patient risk factors, such as medical conditions, that may vary in patient samples. When the distribution of illness differs substantially in patients cared for by different providers, risk adjustment allows for comparison of providers despite these differences.⁴ The intent of each measure specification, therefore, is that each measure should reach its appropriate target population but not over- or under-reach, for such errors in specification waste resources and also may generate misleading conclusions about the quality of care. There are various ways the target population is defined in a measure. Definitions for a population addressed in a measure may be a function of the condition or topic (e.g., concepts that are not disease-specific such as mortality, readmission) that is the focus of the measure, as well as the age, gender, and/or type of healthcare coverage or payer of the individuals eligible for the measure.

Reach of a Measure—Excluded Population

A denominator exclusion or exception criteria can reduce the reach of a measure. A denominator exclusion is an element of an individual measure specification that excludes a population from a measure due to various factors such as age, medical diagnosis, or other descriptive reasons. One example of denominator exclusion for a measure is age, whereby the specification notes that an individual less than 18 years of age is excluded from the measure, thus limiting the denominator to adult patients.

A denominator exception is an allowable reason for excluding a person from a measure despite meeting denominator criteria.⁵ A type of exclusion allowing providers to use clinical judgment to determine whether a patient should be excluded from a measure is also known as an exception. An example of an exception is a patient who would otherwise be eligible for influenza immunization but was exempted from the measure for reasons such as patient preference, medical reason (e.g., allergy), or system reason (e.g., unavailability of vaccine) as documented by the physician. For this study, the research team utilizes the term “exclusion” to refer to both denominator exclusion and exception.

Previous studies have raised complicated issues regarding measure exclusions. Studies have indicated the need to measure quality for excluded populations.^{4,6} Other studies have suggested that exclusions are simply a way to “game” the system and introduce biases that may have ramifications, which include decreasing denominators (thereby inflating the measure performance) and increasing the positive impact on reimbursement in various care settings (see Chapter 6 for information on gaming).^{7,8} Examining measure exclusions to address these complex issues are outside the scope of this study. However, the results of this study may help to inform this debate by providing insights for future measure development and implementation.

Previous studies have examined the exclusions and exceptions in quality measures used in Centers for Medicare & Medicaid Services (CMS) programs. These studies focused primarily on

measures for specific conditions including cardiovascular disease, rheumatoid arthritis, and osteoarthritis measures.^{4, 6, 9-11} Findings from these studies showed that exclusions and exceptions reduced the number of individuals included in the quality measures. Although these exclusions occur for valid reasons, there are a significant number of AMI patients whose care quality on arrival to a hospital may not be measured.

Previous studies have not examined the overall reach of measures used in Medicare programs. This study aims to fill this knowledge gap by examining all Medicare measure specifications in relation to the population characteristics (medical conditions, topics, payer types, etc.) included in each measure. This study also aims to identify patterns of exclusions that limit the reach of Medicare measures.

Objectives

Three objectives guided the analysis to examine populations reached by quality measures used in the Medicare program, and how frequently specified populations are excluded from these measures. These objectives are:

1. Identify and describe the conditions or topics, age categories, and payer types represented in the measures.
2. Identify and describe populations identified in measure exclusions.
3. Identify and describe vulnerable populations excluded from measures.

Methods

Measures and Data Sources

This descriptive study includes 754 current and finalized measures as of December 31, 2013, in 21 CMS Medicare quality measurement programs (Table 5-1). Appendix i-4 contains a hyperlink to a complete list of measures used in this analysis, including National Quality Forum (NQF)-endorsed and non-endorsed measures. The list includes the NQF endorsement status, NQF number if endorsed, and both the measure title used by the CMS program and the measure title used by NQF.

For the analysis, the research team exported each measure name, numerator, denominator, denominator exclusion statements, measure type, NQF endorsement status, measure status, and condition from the CMS Measures Inventory.^{xxxvi} The *Federal Register* rules were the primary source used to populate the CMS Measures Inventory. Measures that were currently implemented in CMS programs as of December 31, 2013, and measures finalized in *Federal Register* final rules as of December 31, 2013, for future implementation in CMS programs were included. The research team

^{xxxvi} The CMS Measures Inventory is a repository of measures used by CMS in its various reporting and payment programs. The inventory contains measure information for measures in a particular CMS program, including Measure Title, Description, Numerator, Denominator, and Exclusion Statements; Measure Type; NQF-Endorsement Information; and Measure Steward. The Inventory is publicly available at <http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/CMS-Measures-Inventory.html>.

crosschecked the exported data against measure specifications for all 754 measures. (See Appendix 5-1 for a list of links to Internet sources for the measure specifications used in this study.)

Table 5-1: Programs Included in Analysis

Setting	Program	Abbreviation
 Hospital	Hospital Inpatient Quality Reporting Program	Hospital IQR Program
	Hospital Value-Based Purchasing Program	Hospital VBP Program
	Hospital Readmissions Reduction Program	HRRP
	Hospital-Acquired Condition Reduction Program	HAC Reduction Program
	Medicare and Medicaid Electronic Health Record (EHR) Incentive Program for Eligible Hospitals and Critical Access Hospitals	EHR EH
	Hospital Outpatient Quality Reporting Program	Hospital OQR Program
	Ambulatory Surgical Center Quality Reporting Program	ASCQR Program
	Inpatient Psychiatric Facility Quality Reporting Program	IPFQR Program
	Prospective Payment System-Exempt Cancer Hospital Quality Reporting Program	PCHQR Program
 Ambulatory	Physician Quality Reporting System	PQRS
	Physician Feedback Program	Physician Feedback Program
	Medicare and Medicaid Electronic Health Record (EHR) Incentive Program for Eligible Professionals	EHR EP
	Medicare Shared Savings Program	MSSP
	Medicare Part C (Display or Star Ratings when applicable)	Part C
	Medicare Part D (Display or Star Ratings when applicable)	Part D
 Post-Acute	Nursing Home Quality Initiative	NHQI
	Home Health Quality Reporting Program	HH QRP
	End-Stage Renal Disease Quality Incentive Program	ESRD QIP
	Hospice Quality Reporting Program	HQRP
	Inpatient Rehabilitation Facilities Quality Reporting Program	IRFQR Program
	Long-Term Care Hospitals Quality Reporting Program	LTCHQR Program

The research team developed exclusion criteria at both the measure level and program level to determine which measures were appropriate for analysis. The research team applied the measure-level criteria below:

- ◆ *Exclude measures that focus on assessing organizational attributes.* This criterion ensures the inclusion of measures with patient-related exclusions in the analysis.
- ◆ *Exclude measures no longer used by CMS.* This criterion ensures that the information in the analysis represents current and future measures.

The research team excluded the CMS programs below from the analysis because these programs allow states to modify measure specifications:

- ◆ Core Set of Children's Health Care Quality Measures for Medicaid and CHIP (Medicaid Child).

- ◆ Core Set of Health Care Quality Measures for Adults Enrolled in Medicaid (Medicaid Adult).

For these measure sets, states have the flexibility to modify specifications published in the Medicaid core set technical specifications manuals. State modification of measures introduces variation in how the measure inclusion and exclusion criteria are defined, making a broad statement about the reach of these measures impracticable.

The research team excluded two additional programs: (1) the Medicare Electronic Prescribing Incentive Program (eRx Incentive Program) because it consists of one measure, which assesses the adoption and use of a qualifying electronic prescribing system, and (2) Physician Compare because measures were not available on the website as of December 31, 2013. The measures used for Physician Compare are measures from programs such as the Physician Quality Reporting System (PQRS) and the Medicare Shared Savings Program (MSSP) that are included in the study.

Analyses

Objective 1—Identify and Describe the Conditions or Topics, Age Categories, and Payer Types Represented in the Measures

To describe the reach of the measures in Medicare programs, the research team examined the number and proportion of measures that reached or included individuals according to the type of payer, condition or topic, and age.

Payer

The research team determined the payer type for individuals included in each measure based on data collection requirements of the program using the measure. Most nursing home measures include data from assessments of all residents in Medicare/Medicaid-certified beds, regardless of payer source. In contrast, some inpatient hospital measures are generated from Medicare claims data and are limited to Medicare patients whose payment for care is reimbursed through the Medicare FFS program.

Condition or Topic Categories

The research team assigned each measure one or more conditions or topics and sub-conditions or subtopics using the taxonomy of measure attributes developed for the U.S. Department of Health and Human Services (HHS) Measure Inventory.¹² Condition represents a broad medical condition or characteristic (e.g., Diabetes and Cardiovascular) and topic represents a non-medical condition or characteristic (e.g., Community Care Coordination/Transitions of Care and Screening) that the measure is assessing. Sub-condition and subtopic represent associated medical (e.g., Nephropathy and Heart Failure) and non-medical (e.g., Medication Reconciliation) conditions or characteristics that the measure is evaluating. The research team classified each measure according to the measure concept by using a taxonomy that includes 40 conditions or topics and 143 sub-conditions or subtopics. A measure may have multiple concepts and thus may be tagged with more than one condition or topic and sub-condition or subtopic. For example, a measure called “Diabetes: Retinal Exam” is classified into the diabetes and

eyes/vision condition or topic categories. The measure is also assigned a subcategory (i.e., eye care) under the diabetes condition category.

The HHS taxonomy was mapped with the list of Top 20 High-Impact Medicare Conditions related to cost, prevalence, variability, improvability, and disparities, as well as level of burden to patients and families.¹³ Table 5-2 maps the Top 20 High-Impact Medicare Conditions to the HHS Inventory taxonomy of conditions; the research team generated sub-conditions from this list.

Table 5-2: HHS Inventory and Top 20 High-Impact Medicare Condition Crosswalk

HHS Inventory Condition	Top 20 High-Impact Medicare Conditions
Cardiovascular	2. HF (Heart Failure) 3. Ischemic Heart Disease 9. AMI (Acute Myocardial Infarction) 15. Atrial Fibrillation
Cancer	7. Breast Cancer 10. Colorectal Cancer 13. Prostate Cancer 16. Lung Cancer 20. Endometrial Cancer
Respiratory	8. Chronic Obstructive Pulmonary Disease
Cerebrovascular	5. Stroke/TIA (Transient Ischemic Attack)
Eyes/Vision	17. Cataract 19. Glaucoma
Mental Health Care & Substance Related Care	1. Major Depression 6. Alzheimer's Disease
Diabetes	4. Diabetes
Renal & Genitourinary	12. Chronic Renal Disease
Musculoskeletal	11. Hip/Pelvic Fracture 14. Rheumatoid Arthritis/Osteoarthritis 18. Osteoporosis

Age

The research team examined both measure denominator and measure exclusion criteria to identify information for these age groupings. For instance, the measure denominator may stipulate that the measure is applicable to patients who are 18 years of age or older, while another specification may explicitly detail the age restrictions in the denominator exclusion. Measure age groupings included 65 years of age and over, 18 years of age and over, and under 18 years of age.

Objective 2—Identify and Describe the Populations Identified in Measure Exclusions

Measure specifications generally clearly identify denominator exclusions. Exclusion criteria may be implied in the measure denominator. For example, a measure denominator may include individuals 18 years of age and older, but the denominator exclusion does not clearly exclude individuals less than 18 years of age. With this in mind, the research team categorized both

explicit and implicit measure exclusions based on an examination of both denominator exclusions and description.

Measure Exclusions

To generate categories for measure exclusions used in the analysis, the research team conducted a pilot test using measures in five selected programs: Hospital Inpatient Quality Reporting Program (Hospital IQR Program), MSSP, End-Stage Renal Disease Quality Incentive Program (ESRD QIP), Nursing Home Quality Initiative (NHQI), and Medicare Part C (Display or Star Ratings when applicable). These Medicare programs met the following criteria for inclusion in the pilot test:

- ◆ Programs representing a variety of care settings.
- ◆ Programs that have both newly implemented measures as well as established measures.
- ◆ Programs that have measures that are both NQF-endorsed and non-endorsed.

The research team used the following steps to develop exclusion categories and test categorization of measures during the pilot:

1. Review the denominator and exclusion criteria of measures.
2. Identify and achieve consensus on exclusion categories.
3. Draw a sample of 30–40 measures.
4. Measure inter-rater reliability.
5. Retrain clinical reviewers if inter-rater reliability coefficient is less than 0.7.
6. Re-measure inter-rater reliability until a coefficient of 0.7 or greater is achieved.

The 19 exclusion categories include both medical and non-medical categories such as medical diagnosis/other clinical reasons, treatment time frame, and baseline measurements outside parameters (see Appendix 5-3 for the complete list of exclusion categories and example exclusions for each category). The exclusion categories and definitions are:

- ◆ **Medical Diagnosis/Other Clinical Reasons:** A patient who was excluded due to allergy or contraindications, infections, urgent/emergent medical conditions, as well as specific conditions that would constitute a contraindication for the treatment.
- ◆ **Provider Discretion:** A patient who was eligible for a measure but excluded at the discretion of the provider. In measures that allow physician discretion exclusions, physicians are able to designate that a patient did not receive the services described in the measure because of patient, medical, or system-related reasons. These patient, medical, and system reasons are not specified in the measure. The provider may use clinical judgment when using this type of exclusion.
- ◆ **End-of-Life Care:** A patient who receives hospice care, palliative care, comfort measures, a patient who has expired, or a patient who is terminally ill.
- ◆ **Patient Refusal:** A patient who refuses care or aftercare, the ability to refuse treatment based on religious preference, refusal due to other reasons or reasons not specified, a patient leaving against medical advice, and elopement.

- ◆ **Transfer To or From Other Facility:** A patient who has been transferred to or from or discharged to other care facilities.
- ◆ **Treatment Time Frame:** A patient who was excluded because the care received fell in a particular time or treatment period, such as the hospitalization period, treatment days, and look-back period.
- ◆ **Baseline Outside Parameters:** A patient excluded due to previous medical findings that fall outside a range or parameter specified by the exclusion criteria.
- ◆ **Continuity of Enrollment:** A patient with gaps in healthcare coverage or with a non-reimbursable treatment.
- ◆ **Missing Data:** A patient excluded due to missing records or information.
- ◆ **Procedure-Specific Exclusion:** A patient excluded due to the receipt of a specified medical procedure.
- ◆ **System Reasons:** A patient excluded due to events related to the healthcare facility where the patient was receiving care.
- ◆ **Discontinuation of Care/Other Non-Clinical Services:** A patient excluded due to inability to tolerate care or for another non-specific patient reason.
- ◆ **Clinical Trials:** A patient excluded due to enrollment in a clinical trial.
- ◆ **Psychiatric Diagnosis or Cognitive Impairment:** A patient with psychiatric diagnoses, cognitive impairment, Alzheimer’s disease, dementia, or who lacks the ability to self-report.
- ◆ **Prior Treatment or Occurrence of Complication:** A patient who received a treatment or experienced a complication prior to admission or arrival at a healthcare facility.
- ◆ **Gender:** The sex or gender excludes the patient.
- ◆ **Length of Stay:** A patient who meets specified length of stay parameters.
- ◆ **Age:** Age excludes the patient.
- ◆ **Payer:** A patient’s source of healthcare payer coverage excludes the patient.

The research team used the inter-rater reliability Kappa statistic to measure the agreement between clinical reviewers. Generally, interpretation of agreement based on Kappa values are as follows: 0.4 to 0.59 indicates moderate agreement, 0.6 to 0.79 indicates substantial agreement, and 0.8 and higher indicates outstanding agreement.¹⁴ An item-by-item Kappa score of average to above-average agreement was set at 0.7 or better to establish inter-rater reliability among the clinical reviewers. The research team achieved an inter-rater reliability Kappa coefficient of 0.7 or greater on categorizing measure exclusions.

The research team compared exclusions found between process and outcome measure types. Process of care measures assess the degree to which care providers perform healthcare processes that are evidence-based, achieve desired aims, and avoid processes that result in poor care.¹⁵ Alternatively, outcome measures assess the status of a specific health condition (e.g., illness or death).

Each current unique measure was reviewed for all measure denominator exclusions; each exclusion was counted as an individual observation. For example, if an exclusion of patient refusal of care or service existed in 35 measures, the research team counted it 35 individual

times. Throughout the analysis, the research team used this approach to account for the frequency of individual exclusions across measure programs.

The research team calculated the mean, minimum, and maximum number of exclusions for each measure included in the analysis and examined the percentage of measures by exclusion category as well as exclusion trends.

Objective 3—Identify Vulnerable Populations That May be Excluded From Measures

The research team examined measure exclusion criteria focused on vulnerable subgroups of the CMS population. These subgroups include individuals with high-risk conditions, psychiatric diagnoses or cognitive impairment, and individuals who are receiving end-of-life care. The high-risk condition category was identified from data elements extracted from the measure specifications that include totally dependent, comatose, and non-responsive patients. The psychiatric diagnosis or cognitive impairment and end-of-life care categories were based on the definitions indicated above. The research team assessed the percentage of measures with these exclusions.

Limitations

This study is limited by the approach used to determine the reach of measures. With the exception of payer, which was examined at the program level, this study assessed the reach of measures by examining the included and excluded populations identified from measure specifications. Other factors such as program-specific participation requirements affect the reach of measures but were not examined in this study. For example, participation in the PQRS was voluntary for eligible professionals until 2013, while programs such as the Nursing Home Quality Initiative (NHQI) require participation of all Medicare-certified nursing homes. Therefore, patients who receive care from non-participating providers or non-Medicare-certified nursing homes are not reached by the measures.

This study used a cross-sectional design and examined measures and their exclusions as specified as of December 31, 2013, for current unique measures used in 21 CMS programs. A longitudinal, year-by-year examination of changes in the measure specifications was beyond the scope of this analysis.

Results

Objective 1—Identify and Describe the Conditions or Topics, Age Categories, and Payer Types Represented in the Measures

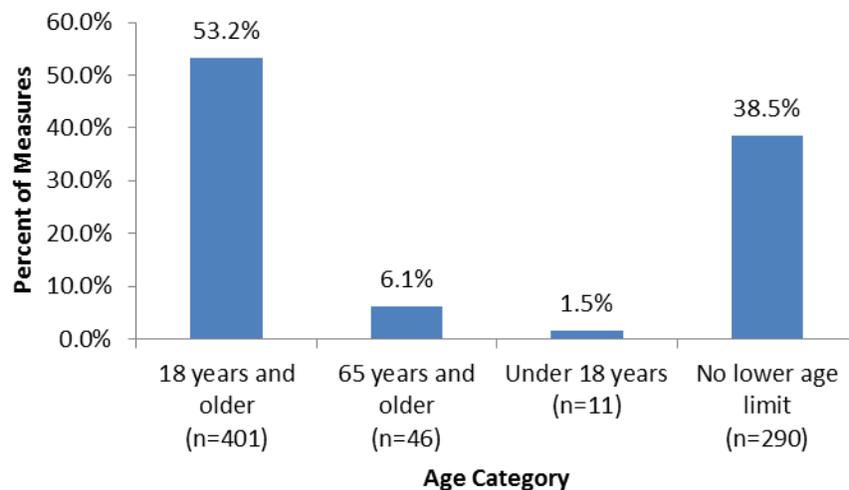
To assess the extent of the reach of the 754 measures across the CMS population, the research team evaluated measure specifications to determine the initial population of Medicare measures. To accomplish this, the research team identified the age requirements, condition or topic for each measure, and payer type.

Age

The age analysis was based on the lower age limit identified in a measure. Figure 5-1 shows the reach of measures used in Medicare programs according to these age limits. Analysis indicated that 6.1 percent of the measures are limited to individuals 65 years of age and older. Analysis also identified that 53.2 percent of the measures are limited to individuals 18 years of age and older, and 1.5 percent of the measures are limited to individuals less than 18 years of age (Figure 5-1). Of the 754 measures, 290 measures (38.5 percent) do not have lower age limits. Six measures (0.01 percent, not shown in graph) had other lower age limits not applicable to age categories included in the graph below.

The age analysis also showed that 653 measures (86.6 percent) do not have upper age limits, and 280 measures (37.1 percent) have neither upper nor lower age restrictions. Additional analysis of exclusion criteria pertaining to age showed that 8.2 percent of the measures exclude individuals over 65 years of age. Therefore, 91.8 percent of the measures include individuals 65 years and over.

Figure 5-1: Percentage of Measures Limited to Age 18 Years and Older, 65 Years and Older, Under 18 Years, and No Lower Age Limit

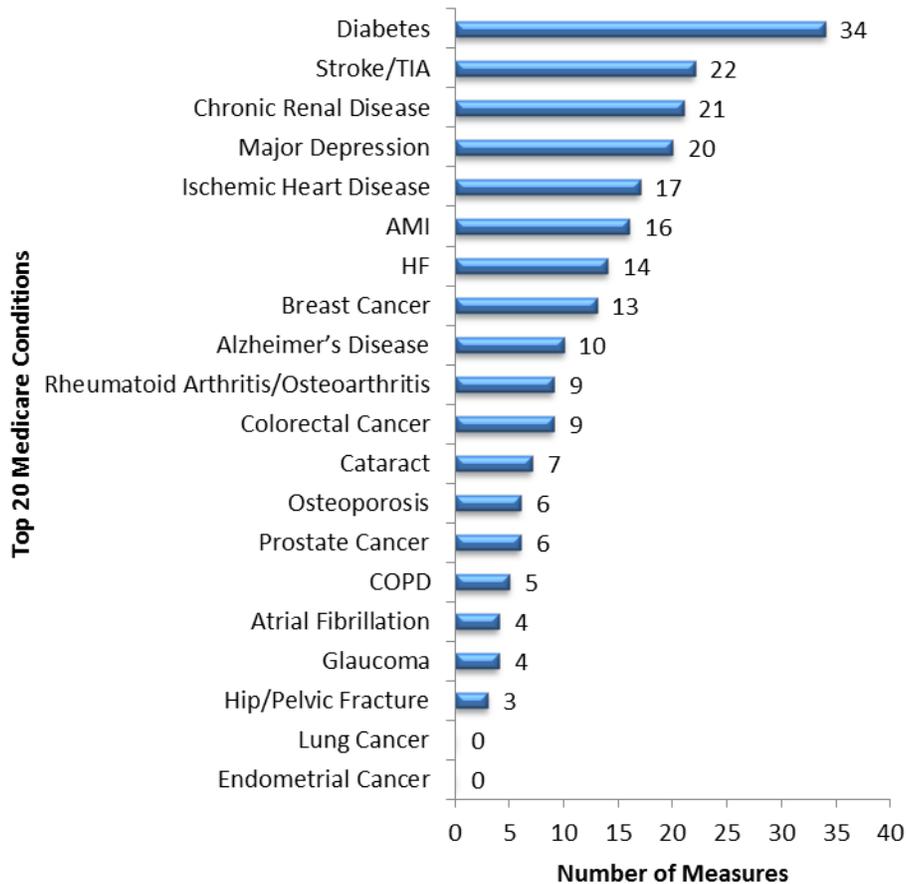


Condition or Topic

Based on the HHS Measure Inventory taxonomy of conditions, 33 conditions or topics and 77 sub-conditions or subtopics were identified across the 754 measures examined. Appendix 5-2 shows the distribution of the measures used in Medicare programs according to condition. The 10 most frequently included conditions or topics in Medicare measures are patient safety, cardiovascular, health services administration, chronic and elder care, mental health and substance abuse-related, functional status, communicable disease, cancer, diabetes, and surgical procedures. Figure 5-2 shows the reach of the measures to patients affected by high-impact conditions. High-impact conditions focused on diabetes, stroke/transient ischemic attack (TIA), chronic renal disease, major depression, ischemic heart disease, acute myocardial infarction (AMI), heart failure (HF), breast cancer, and Alzheimer's disease are represented in 10 or more

Medicare measures. As demonstrated in Figure 5-2, two high-impact Medicare conditions (i.e., endometrial cancer and lung cancer) are not specifically addressed by the measures used in Medicare programs.

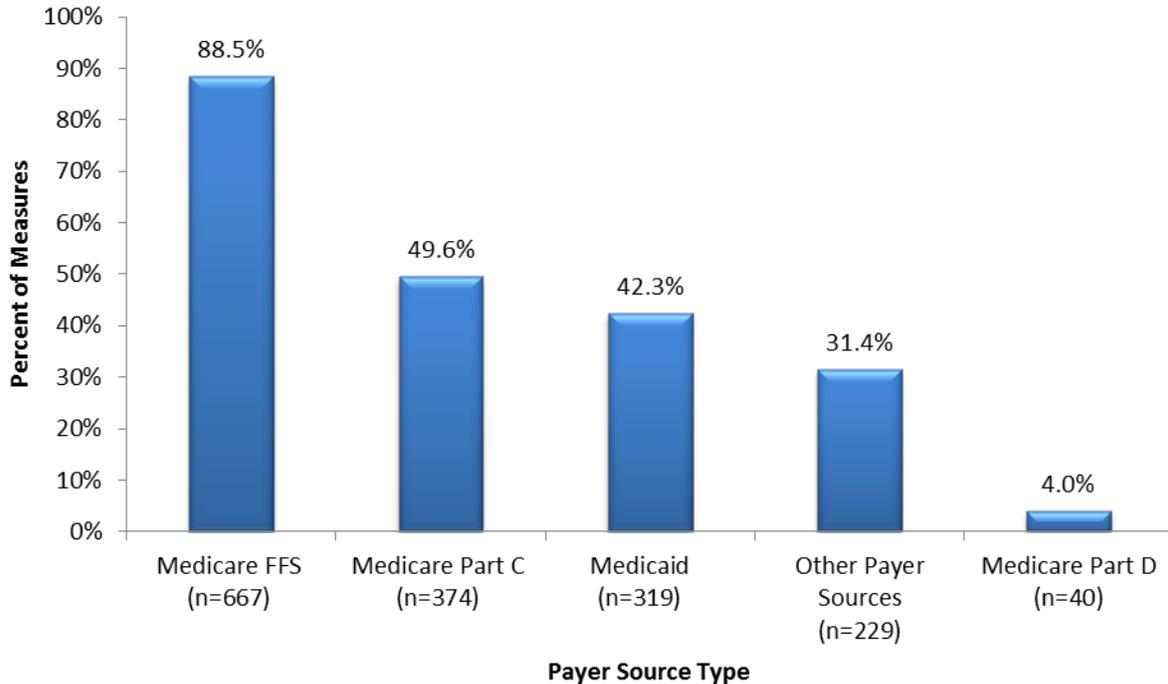
Figure 5-2: Number of Measures That Address High-Impact Medicare Conditions



Payer

Measures may be applicable to one or more payer types. The research team counted each payer type applicable to a measure as a discrete occurrence. Findings showed that 88.5 percent of the measures reviewed included Medicare FFS patients. The percentages of measures that include Medicare Part C and Medicaid patients were 49.6 percent and 42.3 percent, respectively (Figure 5-3). A smaller percentage of measures were found to include patients with other payer sources (31.4 percent) and Medicare Part D (4 percent) payer types.

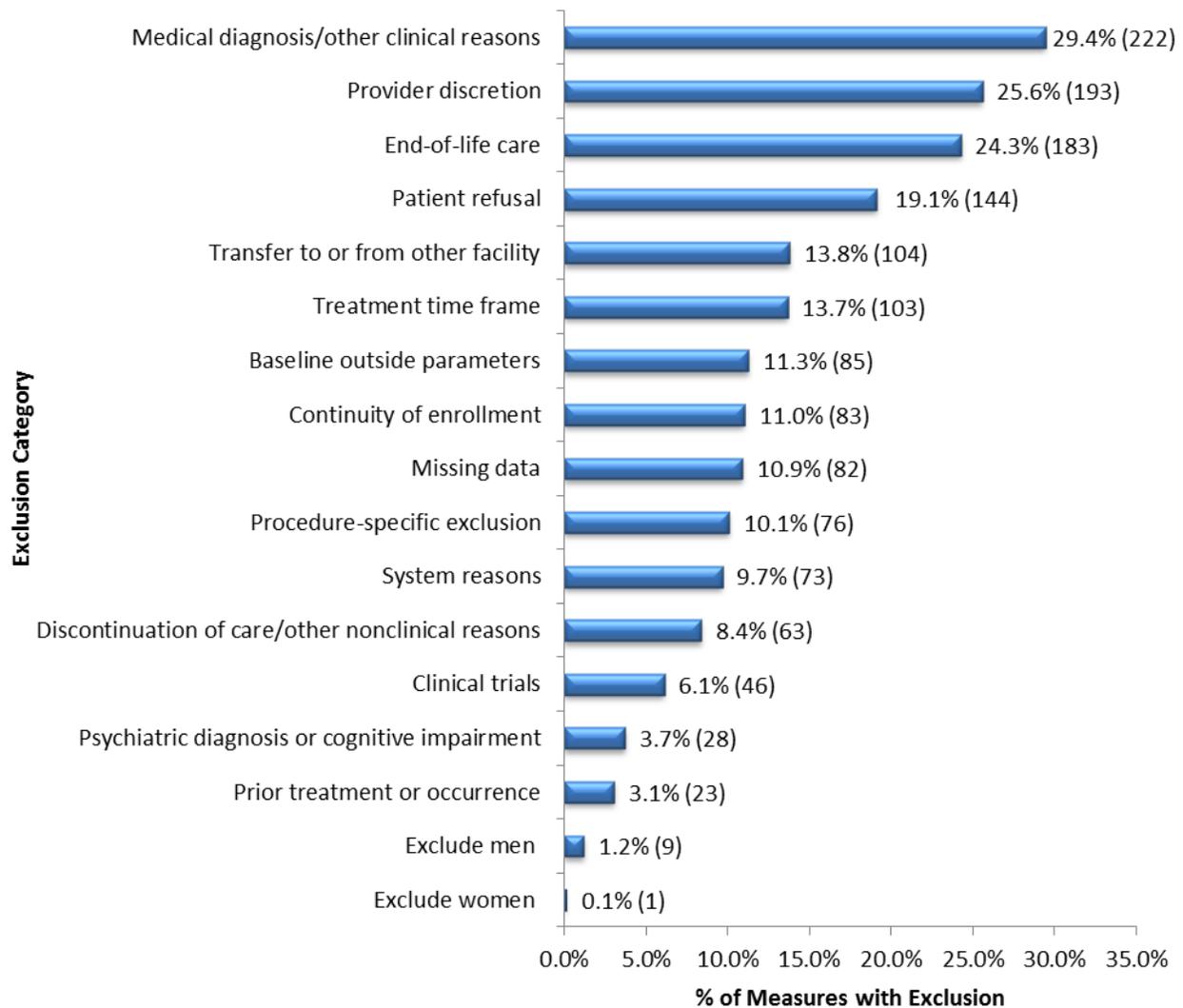
Figure 5-3: Percent of Medicare Program Measures That Include Patients With Medicare FFS, Part C, Part D, Medicaid, and Other Payer Sources (n=754)



Objective 2—Identify and Describe Populations Identified in Measure Exclusions

To understand if patients with specific characteristics were being excluded from measures used in Medicare programs, the research team identified and categorized the exclusion criteria used in 754 measures. The number of exclusions per measure ranged from zero to 21. Figure 5-4 shows that the most prevalent exclusion category^{xxxvii} was medical diagnosis/other clinical reasons (29.4 percent). Exclusion by physician discretion was the second most prevalent (25.6 percent). Gender was the least frequently found exclusion, occurring in 1.3 percent (men or women) of the exclusions.

^{xxxvii} Exclusions based on payer, age, and length of stay are not included in this graph. Data for exclusions based on gender were separated. Individual data for men and women are presented.

Figure 5-4: Exclusion Categories and Rate of Exclusions for All Unique Measures

Frequently Found Measure Exclusion Categories Across Care Settings With 10 or More Measures

The type and number of exclusions varied across programs. While the other analyses in this chapter include each of the programs listed in Table 5-1, the analysis in Table 5-3 is limited to programs with 10 or more measures. Table 5-3 shows the number and frequency for the following five most frequently found exclusions (excluding end-of-life care): (1) medical diagnosis/other clinical reasons, (2) provider discretion, (3) patient refusal, (4) transfer to or from other facility, and (5) treatment time frame. Analysis related to the end-of-life care exclusion category is included in the next section.

All five categories applied to measures used in inpatient hospitals, physician practice settings, and inpatient cancer hospitals. Four of the five categories applied to measures used in outpatient

hospitals. Three of the five categories were found in measures used in home health and nursing home settings. Two of the five categories were found in measures used in dialysis facility centers.

Table 5-3: Summary of Frequently Found Exclusion Categories Across Care Settings With 10 or More Measures

Settings	Measure	Medical Diagnosis/ Other Clinical Reasons	Provider Discretion	Patient Refusal	Transfer To or From Other Facility	Treatment Time Frame
	n	n (%)	n (%)	n (%)	n (%)	n (%)
Inpatient Hospitals	92	51 55.4%	30 32.6%	33 35.9%	75 81.5%	29 31.5%
Physician Practice	460	98 21.3%	159 34.6%	100 21.7%	29 6.3%	30 6.5%
Inpatient Cancer Hospitals [Prospective Payment System (PPS)-exempt]	18	11 61.1%	4 22.2%	1 5.6%	7 38.9%	3 16.7%
Outpatient Hospitals	21	8 38.1%	3 14.3%	2 9.5%	0	3 14.3%
Home Health	86	45 52.3%	0	0	1 1.2%	10 11.6%
Nursing Homes	33	6 18.2%	0	2 6.1%	0	15 45.5%
Dialysis Facility Centers	15	4 26.7%	0	0	0	9 60%

Medical Diagnosis/Other Clinical Reasons

The research team observed the percentage of exclusions due to a specific medical diagnosis or other clinical reason in 50 percent or more of the measures used in inpatient cancer hospital (PPS-exempt) (61.1 percent), inpatient hospital (55.4 percent), and home health (52.3 percent) (Table 5-3). Conversely, medical diagnosis or clinical-related exclusions were present in less than 50 percent of the measures used in outpatient hospitals (38.1 percent), dialysis facility centers (26.7 percent), physician practice (21.3 percent), and nursing homes (18.2 percent).

Provider Discretion Exclusions

Exclusions allowing provider discretion were present in 196 measures used in inpatient cancer hospital (PPS-exempt), inpatient hospital, outpatient hospital, and physician practice settings (Table 5-3). Of these measures, 159 (81.1 percent) were in physician quality reporting programs in the ambulatory setting.

Patient Refusal Exclusions

Analysis revealed that measures used to evaluate quality in inpatient hospital and ambulatory settings excluded patients who refused a treatment or intervention (Table 5-3). Of the 92 inpatient hospital and 460 physician practice measures evaluated, 33 (35.9 percent) and 100 (21.7 percent), respectively, excluded patients who refused a treatment or intervention.

Transfer To or From Another Facility Exclusions

The research team found the highest percentages of measures with this exclusion were reported by inpatient hospitals and cancer hospitals (PPS-exempt). As shown in Table 5-3, 75 (81.5 percent) of the 92 measures implemented in inpatient hospital settings contained these exclusion criteria. Seven (38.9 percent) of the 18 measures used in cancer hospitals were also observed to use transfer-related exclusions. Patients were not included in the eligible population as defined by the measures if they were transferred to care facilities, such as another hospital, or transferred from care facilities, such as an ambulatory surgery center or emergency department/observation unit of another hospital.

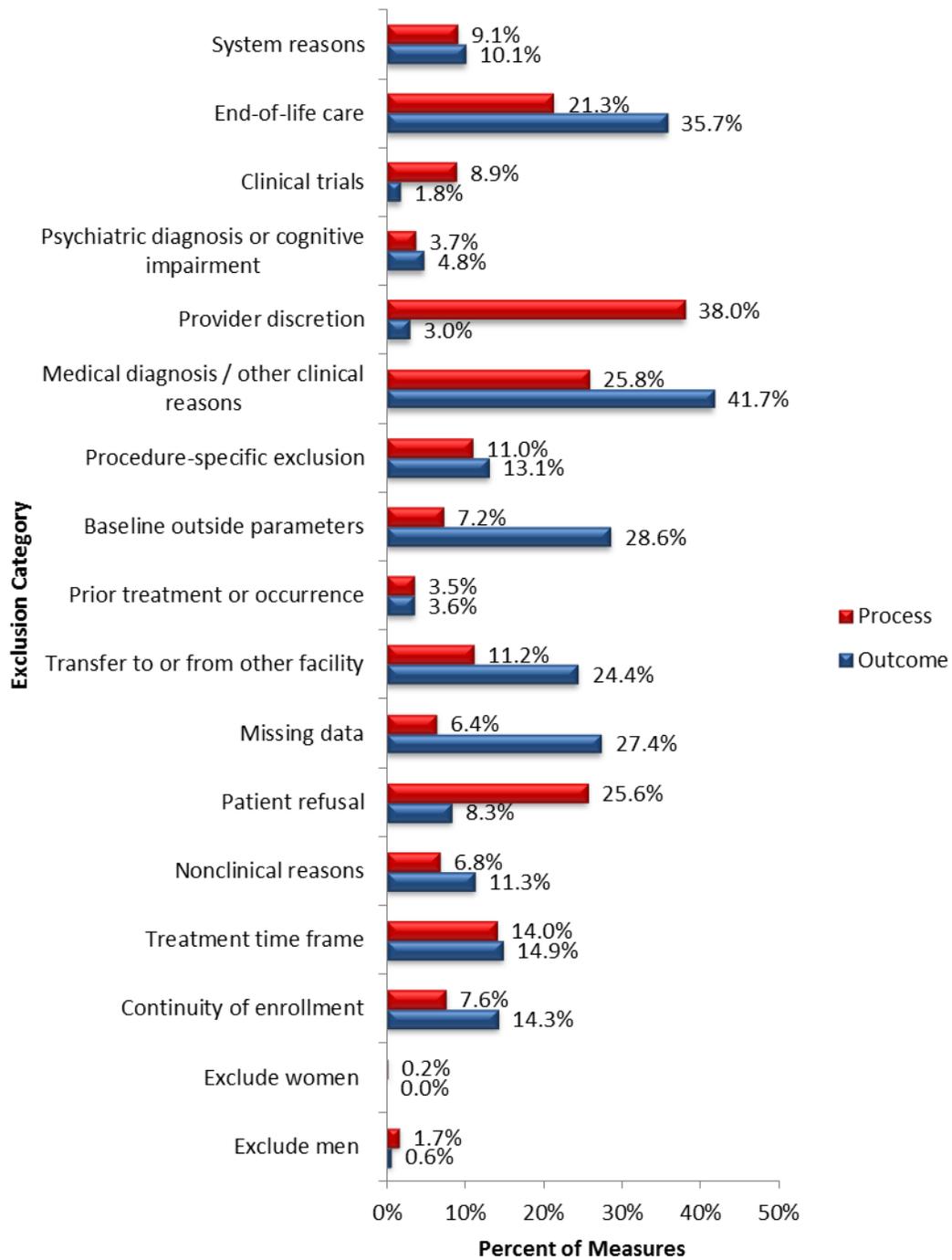
Treatment Time Frame Exclusions

Measures used in dialysis facilities, inpatient hospitals, cancer hospitals (PPS-exempt), outpatient hospitals, nursing homes, and the home health setting were noted to allow treatment time frame exclusions. As shown in Table 5-3, the highest percentages of measures that exclude patients meeting this criterion were observed in dialysis facilities (60 percent), nursing homes (45.5 percent), and inpatient hospitals (31.5 percent).

Measure Exclusions by Measure Types

The research team observed 409 discrete exclusions in 168 outcome measures reviewed, compared to 978 exclusions found in 484 process measures. Compared to process measures, outcome measures had higher percentages of exclusions due to end-of-life care (35.7 percent), medical diagnosis (41.7 percent), procedure-specific exclusion (13.1 percent), baseline outside parameters (28.6 percent), transfer to or from another care facility (24.4 percent), missing data (27.4 percent), non-clinical reasons (11.3 percent), and continuity of enrollment (14.3 percent) (Figure 5-5). Exclusions that allowed for provider discretion (38 percent), patient refusal (25.6 percent), and clinical trials (8.9 percent) were higher in process measures.

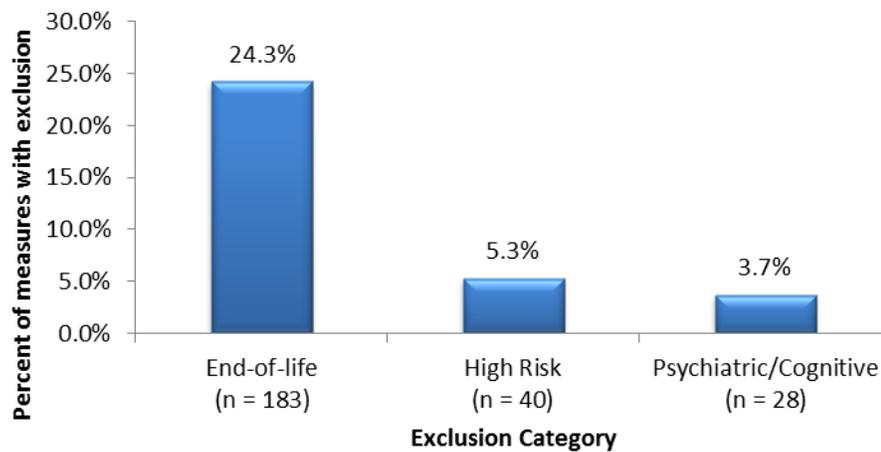
Figure 5-5: Frequency of Exclusion for Process and Outcome Measures



Objective 3—Identify Vulnerable Populations That May be Excluded From Measures

Figure 5-6 depicts the percentage of measures that exclude patients who are undergoing end-of-life care, are high risk, and have psychiatric diagnosis or cognitive impairment. Among the measures examined, 24.3 percent contain end-of-life exclusions; however, 5.3 percent and 3.7 percent of the measures have high-risk and psychiatric-related exclusions, respectively.

Figure 5-6: Percent of Vulnerable Populations Excluded by Category



Discussion

Prior studies have provided evidence that measure exclusions limit the representativeness of CMS measures.^{6, 9, 16} Many measures are derived from current clinical guidelines. As a result, measure specifications may be limited by the evidence-based data that underpin these guidelines. This study provides a baseline understanding of the reach of measures used in 21 Medicare programs. The research team conducted an analysis of measure specifications used in Medicare programs to identify the intended population for the measures. Further, the research team performed an evaluation of criteria to identify populations not represented in or excluded from these quality measures.

The research team examined the reach of measures used by CMS according to the condition or topic, age, and payer type that define the intended population of the measures, as reassigned by the research team for the purposes of this study. Findings showed that many of the quality measures used in Medicare programs reached both Medicare and non-Medicare populations. This finding was reflected in the number of conditions or topics that are the focus of the measures, as well as in the number of measures that include individuals younger than 65. Findings also showed that CMS measures reach populations affected by high-impact conditions including cardiovascular, cerebrovascular, diabetes, musculoskeletal, depression, and Alzheimer's disease. Analysis of the measures demonstrated that many of the measures

encompass individuals with Medicaid and other payer sources. However, there are measures that are limited to a portion of the Medicare population, which is influenced by the level at which the measures are specified (plan versus provider) and by the data source used. Many claims-based measures are limited to Medicare FFS beneficiaries, while measures such as Healthcare Effectiveness Data and Information Set (HEDIS) used in Medicare Part C do not include Medicare FFS beneficiaries. Better alignment of measures sets between these Medicare programs would reduce provider burden and allow healthcare quality to be measured in a comparable manner.

The research team also investigated measure reach by examining measure specifications to identify populations excluded from the measures. Measure developers use exclusion criteria to further define the target population of measures and ensure measure validity. Findings showed that nearly 30 percent of the CMS quality measures examined contained exclusions pertaining to medical diagnoses or other clinical reasons. High rates for this type of exclusion may arise because measures are based on clinical guidelines for the care of patients with certain medical conditions. These findings reflect the latest evidence at the time these measures were developed. While this is a frequently used type of exclusion, the specific criteria vary according to the evidence supporting the measure.

The study results reflected concerns mentioned in previous studies regarding the impact of exclusions in the quality of care for excluded patients.^{6, 9, 16} Studies on the quality of care for Medicare patients with AMI indicated that the percentage of patients who are automatically excluded from AMI process of care quality measures due to medical contraindications has increased significantly over time and has reduced the percentage of ideal candidates for these measures.^{6, 9, 16} Likewise, McCabe et al. examined the impact of exclusions associated with the CMS measure evaluating hospitals' performance in providing primary percutaneous coronary intervention (PCI) within 90 minutes of hospital arrival for an ST-segment elevation myocardial infarction (STEMI).⁴ The results showed that excluded patients tended to have more pre-existing conditions, such as hypertension, heart failure, and peripheral artery disease than included patients. Although medical diagnosis was examined as an exclusion criterion, this study did not identify specific medical conditions or whether individuals with multiple chronic conditions were excluded from quality measures. Given that over two-thirds of Medicare beneficiaries have two or more chronic conditions,¹⁷ further study may be warranted.

Analysis of exclusion criteria also showed that a percentage of individuals might not be represented in measures due to reasons other than automatic exclusions. This is the result of exclusions at the discretion of care providers. The research team found these exclusions in 25.6 percent of all unique measures in the data and a majority of the measures used to evaluate quality of care provided by physicians and other clinicians.

The effects of these discretionary exclusions used in CMS quality measurement programs are not well studied, and little is known about the actual usage rates of discretionary exclusions. Studies that examine the appropriateness of physician exception reporting showed that inappropriate use of exception reporting is infrequent.^{11, 18} A recent study pertaining to physician discretionary exclusions for the PQRS program measures on rheumatoid arthritis and osteoporosis care showed that 5 to 46 percent of patients who visited physicians for these conditions did not receive recommended care for reasons not specified by their physicians.¹⁰ In this study, the

research team validated prescription medication use for patients with osteoporosis and rheumatoid arthritis. Using pharmacy data, the research team found medication prescribed in 49 percent of osteoporosis and 67 percent of rheumatoid arthritis patients whose physicians reported to the PQRS program that the medication was not prescribed.

Studies of the physician pay-for-performance system in the United Kingdom offer insight on the issue of physician exception reporting to maximize incentive payments, albeit the results are mixed. These studies provide information on factors associated with using this type of exclusion and the overall rate of physician use of exception reporting. Studies consistently found that the overall rate of exception reporting was low.^{7, 8, 19} However, findings on the use of exception reporting for financial gain were inconsistent. Doran et al. found no association between the rate of exception reporting and financial gain and noted that practices achieved their clinical targets before taking exception reporting into account.¹⁹ Other studies found that practices that did not receive maximum reimbursement in the previous year had higher exception reporting rates the following year.^{7, 20} Factors associated with exception reporting included characteristics of the physician practice location and practice size. Findings indicated that although the effects were small, higher rates of exception reporting were associated with large practice sizes and with practices located in less affluent areas. Issues pertaining to the inappropriate use of measure exclusions to “game the data” identified in published literature are also discussed in Chapter 6—Measure Use: Unintended Consequences in Hospitals, Nursing Homes, and Ambulatory Settings.

There are advantages to using physician exception (discretionary exclusions) in quality measures. Studies on care approaches for patients with multiple chronic conditions can be used to understand the positive impact of exclusions based on physician discretion. A qualitative study on clinicians’ treatment approaches to older patients with chronic multiple conditions described physicians’ beliefs in tailoring therapy and in the benefits and harms of guideline-directed care.²¹ Physicians acknowledged that quality measures are barriers to identifying the best decision for these patients. Provider discretion exclusions may therefore be a useful approach for physicians to tailor care for patients with multiple chronic conditions and also for patients who may be nearing end of life.

Additional findings from the analysis of measure exclusion criteria revealed exclusions related to end-of-life care and patient refusal in 24.3 percent and 19.1 percent of the measures, respectively. Although these findings seem to indicate that a percentage of measures did not reach these patients, they can also be interpreted as a percentage of measures that are sensitive to the care preferences of patients. Provided that patients are given adequate information with which to make decisions, exclusions based on end-of-life care and patient-specific reasons may ensure that patient autonomy is maintained. Further, these exclusions may safeguard patients from a known pitfall of performance measurement: Treating the measure, not the patient. While a number of measures exclude patients receiving end-of-life care, measures that specifically address the needs of this population are beginning to be developed and used in the recently established Hospice Quality Reporting Program (HQRP).

The research team also examined measure exclusion categories across all Medicare programs to evaluate the impact of exclusions to Medicare subpopulations. Specifically, exclusions that target Medicare patients whose source of vulnerability is due to advanced age, end of life, or

mental or cognitive conditions were assessed. Findings demonstrated that a small percentage of Medicare measures were observed to exclude psychiatric diagnosis or cognitive impairment, those who were comatose, or those who were entirely dependent for care. These exclusions are not found in nursing home measures.

A comparison of process and outcome measures showed higher percentages of provider discretion exclusions in process measures (38 percent) than in outcome measures (3 percent). There are differing opinions among measure development organizations regarding the appropriateness of using exclusions (sometimes referred to as exceptions) not explicitly defined in the measure specifications.

Conclusions

Key Findings

CMS programs and measures reach a wide range of patients with high-impact conditions. CMS quality measures reach 18 of the top 20 high-impact Medicare conditions experienced by beneficiaries; however, measures addressing these high-impact conditions are not evenly distributed across CMS reporting programs.

The study also concluded that the reach of measures used in Medicare programs extends beyond the Medicare beneficiary population and includes individuals with Medicaid or other payer sources. These findings suggest that efforts by CMS to achieve balance in its quality measures and to maximize the representation of individuals reached by the measures are having the intended effect.

The top five exclusions found in Medicare measures focused on patients with certain medical diagnoses or clinical reasons, patients who were excluded at the discretion of their providers, patients receiving end-of-life care, patients who refuse care, and patients who were transferred to or from a facility. The study also found that patients transferred to or from a facility were excluded in over 75 percent of inpatient hospital measures.

No populations were consistently excluded from measures across Medicare programs. Provider discretion was allowed as an exclusion in over one-fourth of the measures in the study, primarily occurring in physician practice measures. Although physician discretion has the potential to allow for “gaming” of measures,^{7, 20} it is an important exception for accommodating patient preferences and needs, thus supporting patient-centered care.

Actions to Consider

CMS programs and measures reach a wide range of patients with high-impact conditions. CMS quality measures reach a large majority of the top 20 high-impact Medicare conditions experienced by beneficiaries; however, measures addressing these high-impact conditions are not evenly distributed across CMS reporting programs.

Action to Consider: Evaluate existing measures addressing high-impact conditions across measure domains. While many measures address certain high-impact conditions

(e.g., cardiovascular disease), opportunities exist to evaluate the effectiveness of these measures to determine if they continue to address CMS goals and assess patient-centered outcomes.

Action to Consider: Prioritize the high-impact conditions and develop a core set of measures across the continuum of care addressing the relevant measure domains. The core set of measures should include patient-centered outcome measures for each high-impact condition. This core set of measures will represent a patient centered, longitudinal, cross-setting set of measures.

CMS quality measures impact patients beyond the Medicare population. Over 40 percent of the measures used in CMS quality reporting programs include individuals whose healthcare is provided by Medicaid and over 30 percent include individuals with other payer sources.

Action to Consider: With new measure development, emphasize data sources, such as electronic health records and all-payer database, to ensure measures have the widest reach across payers and populations.

Patients transferred to or from a facility are frequently excluded from measures for inpatient hospitals. Patient transfers are a common occurrence, and these exclusions result in many patients not having their quality of care assessed.

Action to Consider: Consider developing paired measures for the transferring and receiving facilities to ensure that the quality of care is measured for transferred patients. For example, when developing hospital-setting measures that exclude patients transferred into a facility from another hospital, a measure addressing the same aspect of care can simultaneously be developed for the transferring facility. Alternatively, a “shared accountability” approach attributing a single measure to these facilities can be explored.

Exclusions that allow provider discretion are frequently found in measures for physician practices. The extent, variability, and appropriateness of their use have not been extensively examined.

Action to Consider: Explore the impact of measure exclusions and the effect of measure use on the proportion of the eligible population. Measure developers can conduct the analysis during new measure development and comprehensive review to determine if the measure can effectively impact population health.

Action to Consider: Develop guidance concerning the use of measure exclusions in collaboration with the NQF that aligns with new measure development and comprehensive review. Guidance would provide clear criteria about provider discretion exclusions and would align the use of exclusions in measure specifications for optimal usability.

Reference List

- (1) Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *Am J Public Health*. 1999;89(9):1322-1327.
- (2) Health Services Advisory Group. A Blueprint for the CMS Measures Management System, Version 9.1. Health Services Advisory Group. December 1, 2012. Available at: <http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/Downloads/BlueprintVolume1-combined-v91.pdf>. Accessed May 14, 2012.
- (3) National Quality Forum, Measure Evaluation Criteria. National Quality Forum Web site. May 2, 2013. Available at: https://www.qualityforum.org/docs/measure_evaluation_criteria.aspx. Accessed August 20, 2014.
- (4) McCabe JM, Kennedy KF, Eisenhauer AC, et al. Reporting trends and outcomes in ST-segment-elevation myocardial infarction national hospital quality assessment programs. *Circulation*. 2014;129(2):194-202. National Center for Health Statistics. *Health, United States, 2013: With Special Feature on Prescription Drugs*. Atlanta, GA: Centers for Disease Control and Prevention; 2014.
- (5) Centers for Medicare & Medicaid Services. The CMS EHR Incentive Programs: What You Need to Know About the Clinical Quality Measures. August 30, 2011. Available at: http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/downloads/CQM_Webinar_Slides.pdf. Accessed June 12, 2014.
- (6) Bernheim SM, Wang Y, Bradley EH, et al. Who is missing from the measures? Trends in the proportion and treatment of patients potentially excluded from publicly reported quality measures. *Am Heart J*. 2010;160(5):943-950.
- (7) Doran T, Kontopantelis E, Fullwood C, Lester H, Valderas JM, Campbell S. Exempting dissenting patients from pay for performance schemes: retrospective analysis of exception reporting in the UK Quality and Outcomes Framework. *BMJ*. 2012;344:e2405.
- (8) Gravelle H, Sutton M, Ma A. Doctor Behaviour under a Pay for Performance Contract: Treating, Cheating and Case Finding? *Econ J*. 2010;120(542):F129-F156.
- (9) Rathore SS, Wang Y, Radford MJ, Ordin DL, Krumholz HM. Quality of care of Medicare beneficiaries with acute myocardial infarction: who is included in quality improvement measurement? *J Am Geriatr Soc*. 2003;51(4):466-475.

- (10) Curtis JR, Sharma P, Arora T, et al. Physician explanation for apparent gaps in the quality of Rheumatology Care: Results from the US Medicare Physician Quality Reporting System. *Arthritis Care Res (Hoboken)*. February 2013;65(2):235-43.
- (11) Kmetik KS, O'Toole MF, Bossley H, et al. Exceptions to outpatient quality measures for coronary artery disease in electronic health records. *Ann Intern Med*. 2011;154(4):227-234.
- (12) Agency for Healthcare Research and Quality, U.S. Dept of Health and Human Services. HHS Measures Inventory—Template of Measure Attributes. January 1, 2014. Available at: <http://www.qualitymeasures.ahrq.gov/hhs/about/template.aspx>. Accessed August 20, 2014.
- (13) National Quality Forum. *Measure Prioritization Advisory Committee Report: Prioritization of High-Impact Medicare Conditions and Measure Gaps*. Washington DC: National Quality Forum; 2010.
- (14) Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics*. 1977;33(1):159-174.
- (15) Rubin HR, Pronovost P, Diette GB. The advantages and disadvantages of process-based measures of health care quality. *Int J Qual Health Care*. 2001;13(6):469-474.
- (16) Masoudi FA, Foody JM, Havranek EP, et al. Trends in acute myocardial infarction in 4 US states between 1992 and 2001: clinical characteristics, quality of care, and outcomes. *Circulation*. 2006;114(25):2806-2814.
- (17) Centers for Medicare & Medicaid Services. *Chronic Conditions Among Medicare Beneficiaries, Chartbook, 2012 Edition*. Baltimore, MD: Centers for Medicare and Medicaid Services; 2012.
- (18) Persell SD, Dolan NC, Friesema EM, Thompson JA, Kaiser D, Baker DW. Frequency of inappropriate medical exceptions to quality measures. *Ann Intern Med*. 2010;152(4):225-231.
- (19) Doran T, Fullwood C, Reeves D, Gravelle H, Roland M. Exclusion of patients from pay-for-performance targets by English physicians. *N Engl J Med*. 2008;359(3):274-284.
- (20) Doran T, Kontopantelis E, Fullwood C, Lester H, Valderas JM, Campbell S. Exempting dissenting patients from pay for performance schemes: retrospective analysis of exception reporting in the UK Quality and Outcomes Framework. *BMJ*. 2012;344:e2405.
- (21) Fried TR, Tinetti ME, Iannone L. Primary care clinicians' experiences with treatment decision making for older persons with multiple conditions. *Arch Intern Med*. 2011;171(1):75-80.

Chapter 6—Measure Use: Unintended Consequences in Hospitals, Nursing Homes, and Ambulatory Settings



Question on Implementation

Has the implementation of quality measures been associated with unintended consequences?

Abstract

Background: The Centers for Medicare & Medicaid Services (CMS) uses quality measures to promote better-quality care for Medicare beneficiaries. This study aims to assess the empirical evidence on whether the use of measures had resulted in undesired effects, commonly referred to as “unintended consequences.” The review addressed five types of undesired effects: (1) worsening quality in unmeasured areas of care (“teaching-to-the-test”), (2) providing overtreatment or unnecessary care, (3) reporting inaccurately high performance (“gaming of the data”), (4) avoiding high-risk or challenging patients (“cherry-picking”), and (5) worsening disparities in care. The study team used the findings from this review to inform the development of provider surveys that CMS will conduct as part of the 2018 Impact Report; the surveys will generate national estimates on the type and prevalence of unintended consequences associated with use of CMS quality measures.

Methods: Using search terms for unintended consequences and quality measurement, the research team conducted a systematic review of relevant articles published in PubMed, Cumulative Index to Nursing and Allied Health Literature, EconLit, and grey literature sources such as the CMS reports between 2000 and 2013. Studies published in English and that used strong evaluation design methods, such as randomized trials, cross-sectional evaluations, time series analyses, and controlled before-after study designs were included in the review. After grouping studies by type of unintended consequence and setting of care (hospital, ambulatory clinic, nursing home, or other), the research team assessed the methodological quality of individual studies as good (low risk of bias), fair, or poor (high risk of bias). The research team graded the overall strength of the evidence (high, moderate, low, or insufficient) for the type of hypothesized undesired effect using criteria adapted from

the Evidence-based Practice Center (EPC) guidelines and the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) criteria developed by the Agency for Healthcare Research and Quality.

Results: The search initially retrieved 9,584 citations, of which 9,496 were excluded based on screening of the title or abstract (n=8,593) or full-text of the article (n=903). Among the excluded articles, 7,049 citations did not mention an established CMS quality measure, 2,000 citations used an excluded study design, 442 citations did not examine unintended consequences, and five citations were excluded for other reasons (two articles were not written in English, and three lacked full-text articles). Of the remaining 88 articles, 55 were considered to have poor-quality designs and dropped from review; the evidence synthesis focused on the remaining 33 relevant studies considered to have fair- or good-quality study designs.

Based on the review, there is insufficient evidence of an association between use of quality measures in hospitals and increased prevalence of teaching-to-the-test (zero studies demonstrating negative unintended consequences out of four total fair- or good-quality studies), overtreatment/unnecessary care (zero out of one), or worsening disparities (one out of four). In nursing homes, there is insufficient evidence regarding teaching-to-the-test (zero out of two), cherry-picking (zero out of one), and gaming (zero out of one). In the ambulatory setting, the research team could not identify consistent relationships between use of quality measures and cherry-picking (two out of three), gaming (one out of two), teaching-to-the-test (three out of eight), worsening disparities (one out of four), and overtreatment/unnecessary care (three out of four), leading the research team to assess the evidence as insufficient. However, three studies suggested that intermediate outcome measures of ambulatory care for diabetes may have been associated with overtreatment.

Conclusions: This review finds little empirical evidence to support or refute the possibility that undesired effects have occurred as a result of the use of quality measures. A limited number of high-quality studies have evaluated the relationship between use of quality measures and unintended effects; consequently, the evidence is insufficient to be able to draw conclusions. The few studies that have been conducted show either inconsistent or no relationship between the use of quality measures and unintended consequences; however, future studies may alter these conclusions. Because a limited number of empirical studies have assessed undesired effects, this indicates the challenges of measuring these effects in practice. There remains an ongoing need to monitor for such effects, particularly as CMS evolves its measure programs to incorporate outcome measures and increases the financial risks for providers with poor performance. In addition, CMS can seek to minimize the likelihood of undesired effects through measure design and selection. Intermediate outcome or process measures developed as a balancing pair may mitigate the potential unintended consequences of over- or under-treatment. Guidance can be developed for the appropriate use of exclusions in measure specifications.

Background

The Centers for Medicare & Medicaid Services (CMS), commercial health plans, and state Medicaid programs are using an array of value-based purchasing and public accountability initiatives to incentivize providers to improve healthcare quality and reduce spending.^{1,2} For example, CMS has implemented quality measurement and public reporting programs in numerous clinical settings, including hospitals, ambulatory clinics, home health, and nursing homes to improve clinical care and outcomes and to guide beneficiary choice of provider or health plan. Reporting on the measures by a provider may be voluntary or linked to the use of financial incentives for either reporting of measures (i.e., pay-for-reporting [P4R]) or actual performance (i.e., value-based purchasing or pay-for-performance [P4P]).

Though these performance measurement programs are designed to improve the quality of care and outcomes for patients, the use of quality measures in applications such as value-based purchasing or public reporting³ may inadvertently create incentives for a provider to engage in undesired behaviors to achieve high performance scores. Such undesired behaviors may not result in better-quality care to patients and may even adversely affect patients.⁴⁻¹⁴ Moreover, concerns have also been raised about how value-based purchasing may redistribute resources away from a provider who cares for complex or challenging patients. Given the substantial and ongoing investments that CMS has made in performance measurement and accountability programs, it is important to understand whether these programs have resulted in undesired effects and, if so, to take actions to mitigate them. The most commonly hypothesized effects are summarized below.

Narrowly focusing quality improvement efforts: Performance measurement is designed to induce a provider to focus attention on particular interventions and conditions. However, a provider may then pay less attention to other important areas of care not subject to measurement, potentially reducing quality of care in non-incentivized areas. This is commonly referred to as “teaching-to-the-test.”

Over-treating or inappropriately treating patients: Depending on the design of the performance measurement program, particularly as it relates to the measures used and performance thresholds set to achieve recognition or reward, a provider may seek to improve performance on measures for patients who do not need or would benefit little from the specific care. For example, researchers hypothesized that measuring and holding hospitals accountable for ensuring that patients diagnosed with pneumonia receive antibiotics within set time periods might lead to inappropriate antibiotic use.^{8, 15-17} CMS has twice changed the allowable time period (from four to eight hours and later from eight to six hours) for providing antibiotics in the specifications for the measure in the Hospital IQR Program partly in response to such concerns. Similarly, measurement programs, depending on their reward structure, may incent a provider to work toward high levels of performance even when not appropriate given the patient’s medical condition(s) or preferences (e.g., 99 or 100 percent of patients receive the recommended care process or achieve the outcome). If so, the provider may over-treat patients to meet the performance goal, which may cause undesired side effects. There are patients for whom the recommended treatment may not apply, because of patient preferences for the types of side effects they are willing to tolerate or because the physician is working to balance the treatment of other patient co-morbidities. Although quality measure specifications contain exclusions for patients with certain clinical characteristics, it is not

possible to fully account for all possible types of exclusions. Therefore, expecting a provider to reach 100 percent performance may lead to undesired effects. For example, tight control of diabetes increases the risk of episodes of severe hypoglycemia.¹⁸ Depending on the performance target established for a given measure, a provider could be incentivized to achieve high levels of performance with little or no clinical benefit to the patient. This phenomenon is frequently referred to as “chasing the tail” of the performance distribution.

Gaming the data used to compute performance scores: A provider may subconsciously or intentionally misrepresent patient information when reporting patient data, which would lead to perceived performance gains without actual improvement.¹⁹ Such practices are commonly referred to as “gaming the data,” “upcoding,” or fraud. Payers frequently develop administrative methods, including chart audits, to curb such behaviors, but these tend to be costly to implement. Reporting hospital-acquired conditions (HACs) in administrative claim record fields that are not used to calculate HAC rates is an example of gaming the data. Such actions may have led to underreporting of HACs before algorithms were adjusted to capture additional diagnosis codes on each record.²⁰ Inappropriately using measure exclusions is another form of gaming. Measure specifications generally contain exclusion criteria to remove certain types of patients prior to computing the measure.^{21,22} While patient exclusions are to be expected, a provider may purposefully exclude truly eligible patients when reporting data to be used in calculating quality measures. Another type of gaming occurs when a provider systematically upcodes patient risk factors that are used to predict risk-adjusted outcomes (i.e., the provider represents patients as being sicker than they actually are).

Avoiding challenging patients: A provider may avoid caring for patients who have complex medical conditions or other challenges (e.g., low socioeconomic status, language barriers, or cultural differences) that may make these patients less able to be compliant with the actions the provider recommends. These types of patients may require greater or more varied resources for the provider to achieve high performance on the quality measures, particularly if the measures are outcomes that might be influenced by patient behavior or circumstances. The provider may then try to treat healthier patients (“cherry-picking” or “cream-skimming”) to improve the likelihood of higher performance on quality measures, leading to access problems for less healthy patients.

Worsening disparities in care: Although quality measurement programs aim to standardize care across patient populations, Casalino et al. have hypothesized that incentives might paradoxically worsen racial and socioeconomic disparities in care.²³ Performance incentives could result in a provider with more challenging patient populations, who may perform poorly on quality measures, seeing resources reduced when in fact the provider needs additional resources to invest in systems and care redesign to meet the quality targets.

Objective

The objective of this study is to conduct a systematic review of the published empirical evidence regarding unintended consequences associated with the implementation of quality measures, specifically related to the measures CMS has implemented.

Methods

The systematic review focused on identifying and assessing the published evidence for the association of CMS quality measurement programs with negative unintended consequences. Table 6-1 lists programs and settings that were included in the analysis for this research question.

Table 6-1: Programs Included in Analysis

Setting	Program	Abbreviation
 Hospital	Hospital Inpatient Quality Reporting Program	Hospital IQR Program
	Hospital Value-Based Purchasing Program	Hospital VBP Program
	Hospital Readmissions Reduction Program	HRRP
	Hospital-Acquired Condition Reduction Program	HAC Reduction Program
	Medicare and Medicaid Electronic Health Record (EHR) Incentive Program for Eligible Hospitals and Critical Access Hospitals	EHR EH
	Hospital Outpatient Quality Reporting Program	Hospital OQR Program
	Ambulatory Surgical Center Quality Reporting Program	ASCQR Program
	Inpatient Psychiatric Facility Quality Reporting Program	IPFQR Program
	Prospective Payment System-Exempt Cancer Hospitals Quality Reporting Program	PCHQR Program
 Ambulatory	Physician Quality Reporting System	PQRS
	Medicare Electronic Prescribing Incentive Program	eRx Incentive Program
	Physician Feedback Program	Physician Feedback Program
	Medicare and Medicaid Electronic Health Record (EHR) Incentive Program for Eligible Professionals	EHR EP
	Medicare Shared Savings Program	MSSP
	Physician Compare	Physician Compare
	Medicare Part C (Display or Star Ratings)	Part C
	Medicare Part D (Display or Star Ratings)	Part D
	Core Set of Children’s Health Care Quality Measures for Medicaid and CHIP (Child Core Set)	Medicaid Child
Core Set of Health Care Quality Measures for Adults Enrolled in Medicaid (Medicaid Adult Core Set)	Medicaid Adult	
 Post-Acute	Nursing Home Quality Initiative	NHQI
	Home Health Quality Reporting Program	HH QRP
	End-Stage Renal Disease Quality Incentive Program	ESRD QIP
	Hospice Quality Reporting Program	HQRP
	Inpatient Rehabilitation Facilities Quality Reporting Program	IRFQR Program
	Long-Term Care Hospitals Quality Reporting Program	LTCHQR Program

Data Collection

Using search terms for unintended consequences and performance measurement, the search (conducted November 23, 2013) yielded 9,584 citations published between 2000 and 2013 from PubMed, Cumulative Index to Nursing and Allied Health Literature, EconLit, and grey literature sources such as CMS reports. The research team included published studies that assessed whether CMS quality measures were associated with unintended consequences within hospitals,

ambulatory clinics, nursing homes, or other settings. The research team also included studies of non-CMS quality programs (including non-U.S. programs) if they included measures similar to those used by CMS. The research team used the following search terms when retrieving citations from PubMed:

- ◆ process measures[MH] OR “process of care” OR “process measures” OR “processes of care” OR “process measure” OR ((“NQF” OR “national quality forum”) AND (practices [TIAB] OR measures[TIAB])) OR “process and outcome” OR “process to outcome”
- ◆ “payment for performance” OR “pay for performance”[TIAB] OR p4p[TIAB] OR “pay for value”[TIAB] OR “financial incentive” OR ((bonus[TIAB] OR reward[TIAB]) AND (payment[TIAB] OR reimburse*[TIAB] OR incentive*[TIAB]) AND (quality[TIAB] OR value[TIAB])) OR “quality and outcomes framework”
- ◆ Outcome Measures[MH:noexp] OR Outcome and Process Assessment [mh:noexp]
- ◆ quality indicators, health care[MH:noexp] OR (“quality indicators”[TIAB] OR ((CMS OR Medicare) AND measure*[TIAB]))
- ◆ “quality measures”[TIAB] OR “quality measure”[TIAB]
- ◆ “performance measure”[TIAB] OR “performance measures”[TIAB] OR “process performance” OR “process metrics” OR (“performance-based” AND outcome)
- ◆ “public reporting”[TIAB] OR (public AND “report card”[TIAB]) OR ((Medicare OR CMS) AND (“hospital compare” OR “nursing home compare” OR “Medicare ESRD” OR (renal AND “quality incentive program”) OR “Home health compare”))
- ◆ ((unintended OR unforeseen OR unplanned OR adverse) AND consequences[TIAB]) OR “unintended impacts” OR “unintended effect” OR “negative effect” OR “negative effects” OR “unintended effects” OR pitfall* OR gaming[TIAB] OR “cherry-picking” OR “cherry picking” OR ((withhold OR denial) AND (treatment* OR procedure*)) OR (unmeasured AND (quality OR performance)) OR (avoid* AND high-risk) OR (avoid AND (minority OR minorities))OR (divert AND attention) OR side effect* OR adverse effect* OR ((racial OR underserved OR socioeconomic) AND disparities) OR “cream skimming” OR ((case selection) AND (bias OR denial)) OR (perverse AND incentives) OR (unmeasured AND (care OR area)) OR (explicit* AND link*) OR ((incentivized OR rewarded) AND (unrewarded OR non-incentivized OR nonincentivized)) OR “patient avoidance” OR ((overuse OR unwarranted OR overutilization) AND measure[TIAB]) OR (performance AND selection) OR (“pay-for-performance”) OR “public reporting” OR ((non-incentivized OR nonincentivized OR non-incentivised OR nonincentivised) AND (measure[TIAB] OR indicator[TIAB] or measures[TIAB] or indicators[TIAB])) OR “unintended outcomes” OR “unintended outcome” OR (overtreatment OR “overtreatment”) OR ((deprivation OR inequalities[TIAB]) AND (indicator OR measure OR quality OR incentive)) OR (“exception reporting”) OR (internal motivation OR (professional AND satisfaction) OR professionalism) OR (non-incentivized AND performance AND quality)
- ◆ English[LA] AND 2000:2015[dp] NOT (((letter[pt] OR editorial[pt] OR comment[pt]) NOT clinical trial[pt]) OR (animal NOT (human OR humans)))
- ◆ Final Search (#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7) AND #8 AND #9

Two investigators (including one clinician) independently evaluated each article for inclusion at two stages. In the first stage, the research team screened articles for relevance using the title and abstract after completing a training session using 30 citations. At the second stage, the research team screened full-text articles of those citations passing the first stage. The research team excluded citations that did not describe a strong empirical evaluation approach that mitigated sources of confounding; acceptable study designs included controlled trials, controlled before-after studies, interrupted time-series analyses, and cross-sectional analyses to test the association between use of measures and unintended effects. The research team flagged systematic reviews to use as background material and to extract relevant studies included in the original articles but otherwise did not include them in the results. Finally, articles that did not empirically test whether a quality measure led to unintended consequences in actual practice were excluded; this latter category included simulation studies and studies testing the development of new measures.

Two investigators then independently abstracted data from each study passing the second screening stage. Abstracted data included provider type; number of patients and providers; level of analysis (patient, provider, region); measure type (process of care, outcome, structure of care, efficiency, patient experience); context in which the measure was used (e.g., whether the measure was used by a pay-for-reporting program); type of unintended consequence; and magnitude of effect. The research team followed Cochrane collaboration guidelines to increase the reliability of data abstraction. The research team gave each reviewer a detailed description and examples for each type of unintended consequence and other data elements to be abstracted (such as study design), and resolved disagreements between the original reviewers after discussion among the principal investigators.²⁴

Analysis

Given that there are no established criteria for grading the quality of policy intervention studies, the research team created criteria (discussed in greater detail below) adapted from the Evidence-based Practice Center (EPC) guidelines and the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) criteria developed by the Agency for Healthcare Research and Quality.^{25,26} The research team graded studies in a stepwise fashion. First, to be judged as fair- or good-quality, studies must have demonstrated an adequate method for resolving unmeasured confounding using techniques such as instrumental variables, difference-in-differences or controlled before-after designs, or propensity score matching designs.²⁷⁻²⁹ Studies using time-series data must have statistically addressed underlying trends in outcomes, using interrupted time series or other methods. Second, to be classified as good quality, a study would need to have used well-defined outcomes that were formally adjudicated, previously tested, or were unlikely to be miscoded (death, blood pressure measurements, medication prescriptions, etc.). For example, a fair-quality study might control for confounding but might not use fully validated outcomes or have another deficiency on one or two criteria. Disagreements between investigators were resolved by consensus among the research team. The research team did not measure agreement formally (using a kappa statistic, for example), because inclusiveness rather than perfect agreement was the goal. The research team focused on those articles judged as fair or good quality.

The research team categorized studies by unintended consequence and setting (e.g., studies hypothesizing a link between quality measures and teaching-to-the-test in hospitals). The

research team assessed the strength of the evidence for each hypothesized effect using criteria that were also adapted from the EPC guidelines.^{26,30} However, the original EPC criteria pertained primarily to understanding the effects of interventions that were studied using a controlled design, such as randomized controlled trials (RCTs). Limiting this study solely to RCTs or other controlled studies would have eliminated all published evidence that examined national implementations of quality measurement programs. Therefore, the research team modified these criteria but evaluated the quality, quantity, and consistency of the evidence base in keeping with the original intent of the quality criteria.¹⁰ The research team assigned the studies one of four qualitative strength-of-evidence grades after achieving consensus among all investigators:

- ◆ **High:** There was a high degree of confidence that the evidence reflects the true effect, which is unlikely to change with additional research.
- ◆ **Moderate:** There was a moderate level of confidence that the evidence reflects the true effect.
- ◆ **Low:** There was low confidence that the evidence reflects the true effect and that further evidence is likely to change the estimate. A low rating indicates that there is a high risk of bias and residual confounding.
- ◆ **Insufficient:** There was unanimous consensus that there was a lack of evidence to estimate the effect(s).

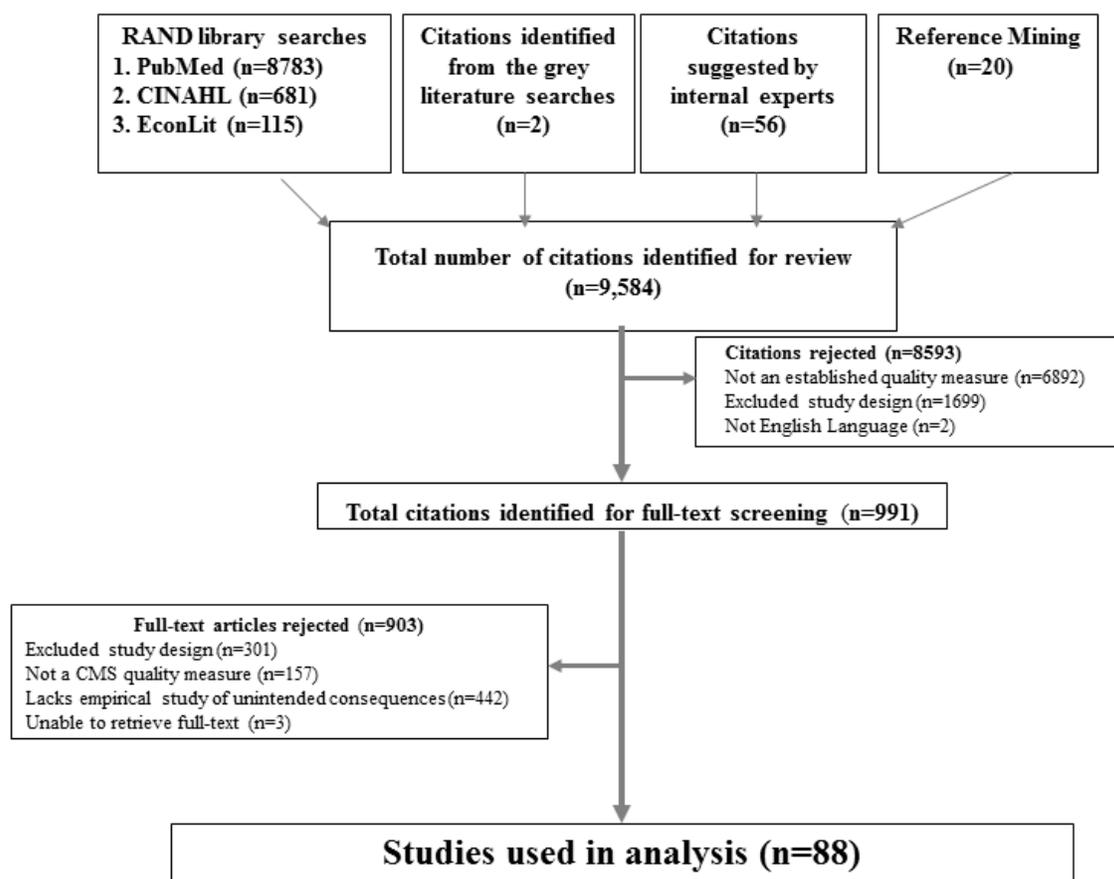
Limitations

This review is subject to several limitations. First, the research team may have missed studies that did not describe consequences as unintended or may not have mentioned consequences; to mitigate this risk, the research team employed the broad search strategy discussed above. Second, strength-of-evidence determinations are inherently subjective in a narrative review, which may bias the results. To reduce the risk of bias, the research team used several experienced clinicians and policy researchers to evaluate the final studies and arrive at a consensus on each question. Finally, the association between performance measurement programs and undesired effects depends on the types of measures used, the size and nature of the incentive target (e.g., bonus or reduction in payment, goal of 100 percent or <100 percent performance), and the setting (e.g., hospital or nursing home). The research team divided the analysis by category to facilitate comparisons across such different types of characteristics. However, the paucity of empirical studies limited the ability to detect whether particular incentive features or measure types influence the prevalence of unintended consequences; this limitation is unavoidable when conducting literature reviews of topics with relatively few articles. Studies aiming to detect gaming or miscoding frequently require costly chart reviews, which may act as a substantial barrier to increasing the evidence base.

Results

Figure 6-1 shows the counts of citations retrieved at each stage. Articles were retrieved predominantly from PubMed. The vast majority (9,496 out of 9,584) were not relevant to the analysis. After summing across both screening stages, article exclusions belonged to four categories: 7,049 citations did not mention an established CMS quality measure (6,892 citations lacked a quality measure, while 157 did not discuss CMS quality measures); 2,000 citations used an excluded study design (1,699 were excluded during the abstract review, while 301 were excluded after screening the full-text); 442 articles did not examine unintended consequences; and five citations were excluded for other reasons (two articles were not written in English, and three lacked full-text articles). Of note, 17 systematic reviews were among the citations excluded because of study design, but all relevant references were extracted from these articles and incorporated into the review within the “Reference Mining” category.

Figure 6-1: Number of Studies Used in Analysis (n=88)



The research team located 88 studies, of which 33 were of fair or good quality. The remaining 55 studies were of poor quality and were not considered in the evidence synthesis. Twelve of the 33 fair- or good-quality studies supported the hypothesis that quality measurement leads to

unintended consequences. Table 6-2 provides the detail counts by setting and topic of the number of studies used in the analysis, and those that were of fair or good quality. There were few citations in each topic area; for example, the most frequently studied area was of unintended consequences in the ambulatory setting, which was the focus of eight fair- or good-quality studies. Results across studies were frequently inconsistent, and many studies found small effects.

Table 6-2: Counts of Studies Related to Unintended Consequences^{xxxviii,xxxix}

Topic	Total Number of Articles per Setting (fair or good quality in parentheses)			
	Hospital	Ambulatory	Nursing Home	Other
Teaching-to-the-test	5 (4)	17 (8)	2 (2)	0 (0)
Gaming	6 (0)	15 (2)	2 (0)	1 (0)
Overtreatment/unnecessary care	3 (1)	4 (4)	0 (0)	0 (0)
Cherry-picking	0 (0)	4 (3)	2 (2)	0 (0)
Worsening disparities	6 (4)	21 (4)	1 (0)	1 (0)
Total	19 (8)	60 (21)	7 (4)	2 (0)

Unintended Consequences of Quality Measurement in the Hospital Settings

The research team identified 19 relevant studies of unintended consequences of U.S. quality measurement programs in the hospital setting (Table 6-2), of which eight were of fair or good quality. There was insufficient evidence to indicate that quality measurement in hospitals was associated with increased prevalence of teaching-to-the-test (zero studies demonstrating negative unintended consequences out of four total fair- or good-quality studies), overtreatment/unnecessary care (zero out of one), or worsening disparities (one out of four).

Four fair- or good-quality studies found no evidence for teaching-to-the-test.^{15, 31-33} However, the studies reported on a limited number of measure areas, and one study, Glickman et al., compared performance in incentivized measures to performance in non-incentivized measures.³³ Glickman et al., Ryan, and Jha et al. also studied the Premier Hospital Quality Incentive Demonstration (HQID) project and found no significant improvements in several incentivized measures (such as acute myocardial infarction mortality), demonstrating that there is no evidence of teaching-to-the-test.^{31, 32} Given these limitations, the research team assessed the evidence for teaching-to-the-test in the hospital setting to be insufficient.

Six poor-quality studies addressed hospitals gaming the data to improve performance scores.^{20, 34-38} The research team assessed the evidence for hospital gaming in response to public reporting (or value-based purchasing) to be insufficient because the six studies the research team identified were poor-quality studies (and potentially biased). These studies indicated that there was substantial scope of gaming (as evidenced by changes in coding over time or underreported

^{xxxviii} Totals are not necessarily the sum of columns because the study may be counted in multiple rows. For example, a study may address teaching-to-the-test and overtreatment.

^{xxxix} As explained in detail in the methods, numerous studies were judged to be of poor quality, often because the studies noted an association but did not adequately account for alternative factors that might have explained the finding. The research team included studies of good or fair quality in the evidence synthesis (i.e., the counts in parentheses in the table).

adverse events), but these studies did not mitigate confounding factors that could explain both the CMS quality measurement programs and hospital coding changes, such as general changes in coding practices. Moreover, studies could not conclusively demonstrate that changes in coding practices were inaccurate.

Three studies addressed hospitals pursuing unnecessary treatments to improve measured performance; each study addressed timing of antibiotics in pneumonia.¹⁵⁻¹⁷ Although there are substantial theoretical concerns for this measure leading to unnecessary care, Friedberg et al. was the one study conducted of fair quality to address this topic. This study found no effect on unnecessary treatments. However, that study used data from 2001–2005, when adherence to the antibiotic timing measure was somewhat lower; repeating a similar study now might yield higher estimates of unnecessary treatment, as might studies on other measures. As a result, the research team judged the empirical evidence for CMS measures leading to unnecessary care to be insufficient.

Four fair- or good-quality studies addressed whether quality measurement programs might have worsening disparities in care for racial minorities or patients from lower socioeconomic groups.³⁹⁻⁴² Werner et al. compared changes in performance in safety-net hospitals (a safety-net hospital is defined as a hospital with a high proportion of Medicaid patients) to changes in performance in hospitals with lower proportions of Medicaid patients after Hospital Compare was implemented (2004 to 2006).⁴² Hospitals with few Medicaid patients (i.e., fewer disadvantaged patients) improved their composite performance by 3.8 percentage points, or 1.5 percentage points more than in safety-net hospitals, suggesting evidence for widening disparities under pay-for-reporting. However, Ryan, Ryan et al. and Jha et al. found that racial or socioeconomic disparities did not widen in hospitals involved in the HQID project (a value-based purchasing demonstration) relative to hospitals under pay-for-reporting.³⁹⁻⁴¹ The research team therefore assessed the strength of evidence for quality measurement causing widening racial disparities to be insufficient.

Unintended Consequences of Quality Measurement in the Ambulatory Setting

Sixty relevant studies pertained to the ambulatory setting, of which 21 were of fair or good quality (Table 6-2). Thirty-six of the 60 studies evaluated quality measurement programs in the UK. Of note, many UK studies examined the effects of ambulatory measures that were similar to those of CMS but were subject to larger financial incentives. The research team did not identify consistent relationships between quality measurement in the ambulatory setting and cherry-picking (two studies demonstrated negative unintended consequences out of three total fair- or good-quality studies), gaming (one out of two), teaching-to-the-test (three out of eight), worsening disparities (one out of four), and overtreatment/unnecessary care (three out of four), leading the research team to assess the evidence as insufficient.

The research team identified 17 studies that addressed teaching-to-the-test, of which eight were deemed to be of fair or good quality.^{7,43-48} The eight fair- or better-quality studies examined the effect of quality measures on performance in non-incentivized areas, under the hypothesis that incentives would lead a provider to focus on incentivized conditions to the detriment of non-incentivized conditions. These studies reached conflicting conclusions while examining a

limited number of non-incentivized conditions and measures. Ganz et al., Mullen et al., Fagan et al., Sutton et al., and Guthrie et al. found no evidence for teaching-to-the-test in either the U.S. or UK performance measurement programs.^{7,45-48} In contrast, large studies by Doran et al., Campbell et al., and Campbell et al. found evidence for teaching-to-the-test in response to the UK Quality and Outcomes Framework (QOF) program.^{43,44,49} Of note, both Campbell et al. and Campbell et al. assessed a similar non-incentivized outcome (continuity-of-care) from the same population, so they were weighted as a single study. Doran et al. estimated that performance on non-incentivized clinical process measures did not actually decline but was 5.6 percent below the predicted performance level. Given the mixed evidence, the research team judged the strength of evidence to be insufficient with respect to whether quality measure programs have induced providers to limit their focus to measured areas at the expense of important unmeasured areas.

The research team identified 15 studies that addressed gaming in the ambulatory setting, of which two were of fair or good quality. In one small fair-quality study, providers reduced their coding of particular diagnoses (International Classification of Diseases 9th edition [ICD-9] diagnostic code 466) for which they would be penalized for inappropriate antibiotic ordering and increased their usage of another diagnostic code (ICD-9 diagnostic code 490) for which they would not be penalized.⁵⁰ Total antibiotics prescribed under both diagnostic codes declined slightly, suggesting that providers changed their coding in response to the rule change to avoid penalties but did not change actual practice. A UK study did not find evidence that providers improved performance by labeling borderline patients as hypertensive, thus increasing the number of easy-to-treat patients.⁵¹ The research team also identified 13 lower-quality studies that examined gaming. Eight studies noted that higher “exception reporting” rates were associated with lower socioeconomic status or higher performance, which suggests that providers may be using exceptions for gaming.⁵²⁻⁵⁹ However, these studies did not address whether QOF induced the providers to over-report exclusion rates. For example, Doran et al. found low rates of exception reporting (2.7 percent) but did not ascertain whether these cases were appropriate.⁵⁸ Therefore, the research team judged the strength of evidence to be insufficient because there are a few inconsistent, fair- or better-quality reports.

The research team identified four fair- or good-quality studies that examined whether quality measurement programs led to unnecessary treatment.⁶⁰⁻⁶³ The results from three of these four studies suggested that certain quality measures may lead to overtreatment when attempting to control hypertension, hyperlipidemia, or hyperglycemia in diabetic patients. Two of the three studies were conducted in Veterans Affairs populations^{62,63} and one was conducted in Taiwan; none used incentives identical to those in the CMS population, so conclusions about the effect of such measures in the CMS population could change with further research. The research team assessed the strength of evidence for intermediate outcome measures for diabetes causing overtreatment to be low. Given the fact that one small but fair-quality RCT suggested no effect in treating patients with hypertension,⁶¹ there is insufficient evidence for concluding the programs were associated with overtreatment.

Three out of the four studies that examined whether primary care practice groups engaged in cherry-picking were of fair or good quality.⁶⁴⁻⁶⁶ The research team identified a single fair-quality study that found no evidence of cherry-picking while studying programs similar to CMS programs.⁶⁵ Two good-quality studies from the Taiwan pay-for-performance program used measures similar to CMS current ambulatory diabetes measures.^{64, 66} However, these may not be

generalizable to CMS programs given that Taiwanese providers were allowed complete discretion over selecting the patients to be used when calculating performance. The research team therefore assessed the strength of evidence for cherry-picking to be insufficient.

The research team identified 21 studies that tested whether quality measurement programs worsened disparities in care between different racial and socioeconomic groups, of which four were deemed to be of fair or good quality. The preponderance of these 21 studies came from the UK, and the studies were inconsistent in showing no increases in disparities. Most studies did not adequately account for underlying trends and were judged as poor quality. One fair-quality study from the UK comparing trends before and after the QOF pay-for-performance program was implemented did show that QOF may have caused (at most) a transient increase in disparities in systolic blood pressure (-5.3 mm Hg in Whites versus -2.3 mm Hg in Blacks).⁶⁷ However, widening disparities were not found in three studies examining a longer time period.⁶⁸⁻⁷⁰ The research team assessed the evidence related to worsening disparities as insufficient, particularly with respect to U.S. quality measurement programs.

Unintended Consequences of Quality Measurement in Nursing Homes

The research team identified seven studies in the nursing home setting; four of the studies were deemed to be of fair or good quality (Table 6-2).⁷¹⁻⁷⁴

In examining unintended consequences of quality measurement in nursing homes, the research team assessed that there was insufficient published evidence for teaching-to-the-test (zero studies demonstrating negative unintended consequences out of two total good- or fair-quality studies), cherry-picking (zero out of two), and gaming (zero out of one).

The research team did not find published evidence of cherry-picking by nursing homes; two higher-quality studies showed no effect⁷² or a small effect.⁷¹ Werner et al. and Arling et al. provided evidence against teaching-to-the-test, but the research team deemed the total evidence to be insufficient.^{73, 74}

Discussion

The research team found limited published evidence to support or refute previously published hypotheses stating that that use of quality measures led to unintended consequences.²³ The systematic review identified 33 studies that were of fair or good study design; these studies examined five types of undesired effects. The higher-quality studies addressed a limited range of clinical areas and the findings were frequently inconsistent (with less than half supporting the presence of unintended consequences and with most others finding small effects). The research team assessed the evidence for use of intermediate outcome measures for diabetes being associated with overtreatment to be low because three fair- and good-quality studies showed consistent potential overtreatment associated with intermediate outcome measures targeting hypertension, hyperlipidemia, and hyperglycemia in diabetic patients. However, the results may not be generalizable to CMS beneficiaries and providers because the studies were conducted on Veterans Affairs medical centers and ambulatory facilities in Taiwan; the incentives used differ from those employed by CMS.^{60, 62, 63} The strength of evidence regarding nursing homes engaging in substantial cherry-picking was deemed to be insufficient because two fair-quality

studies suggested that nursing homes have not been successful in avoiding sicker patients to a substantial degree.^{71, 72} The research team assessed the evidence for the other major categories (teaching-to-the-test, promotion of low-value care, gaming, cherry-picking, and worsening disparities in ambulatory, hospital, nursing home, and other settings) to also be insufficient. These findings are similar to those of prior reviews that studied particular incentive types such as public reporting³ and pay-for-performance.¹⁰

Though many have explored undesired effects of performance measurement programs, there have been many challenges associated with conducting these studies. First, the sponsors of performance measurement programs rarely collect other measures that are not subject to reporting or other incentives to be able to determine if non-measured areas are not improving or declining. Second, data on the clinical and socioeconomic characteristics of patients and the characteristics of the providers are frequently lacking to assess potential negative effects on disparate populations. Third, studies aiming to detect gaming or miscoding regularly require costly chart reviews, which act as a substantial barrier to increasing the evidence base.

Conclusions

Key Finding

This review finds little empirical evidence to support or refute the possibility that undesired effects have occurred as a result of the use of quality measures by CMS between 2000 and 2013. There are a limited number of high-quality studies that have evaluated the relationship between use of quality measures and unintended effects; consequently, the evidence is insufficient to be able to draw conclusions. The few studies that have been conducted show either inconsistent or no relationship between the use of quality measures and unintended consequences; however, future studies may alter these conclusions. The limited number of empirical studies that have assessed undesired effects highlights the challenges of measuring these effects in practice.

Actions to Consider

There remains an ongoing need to proactively monitor for undesired effects, particularly as CMS evolves its measure programs to incorporate outcome measures and increases the financial risks to providers of poor performance. In addition, CMS can seek to minimize the likelihood of undesired effects through measure design and selection.

Auditing efforts can discourage undesired behaviors by a provider. As the size and scope of incentives to a provider increase, so may the incentive for a provider to engage in undesired behaviors to ensure that they meet goals to achieve payments; as a result, efforts to prevent, monitor, and mitigate unintended consequences will remain important.

Action to Consider: Explore development of proactive validation of electronic clinical quality measure/electronic health record data by establishing front-end edits and clinical algorithms appropriate to a given measure. This process could be designed to detect differences in observed versus expected values for associated outcomes or processes, improperly coded risk factors for case mix adjustment, and inaccurate measure exclusions. For example, such a system could use clinically relevant data elements

available in electronic health records (e.g., the expected frequency of renal function testing) to validate particular measures (e.g., appropriate renal dosing). Data that are significantly outside of expected ranges could be flagged and used to target third-party validation specific to each program/setting performed by a CMS contractor or authorized vendor. This process could help to deter providers' gaming of the data used to assess quality and reduce costs associated with traditional auditing.

Action to Consider: Consider establishing a third-party data validation process specific to each program/setting performed by a CMS contractor or authorized vendor to ensure accurate reporting and to provide insight regarding potential unintended consequences of quality measurement. Data collected from this process could enable identification of unintended effects and may allow rapid modifications to measure specifications.

Measure design and selection have the potential to reduce the incentives for a provider to engage in undesired behaviors. For example, an outcome measure penalizing under-treatment (e.g., hyperglycemia) could be paired with a balancing outcome measure penalizing over-treatment (e.g., hypoglycemia).

Action to Consider: Emphasize the development of balancing measures in new measure development projects. Intermediate outcome measures (e.g., glycemic control), or process measures developed as a balancing pair might mitigate the potential unintended consequences of over- or under-treatment.

While patient exclusions are to be expected, a provider may purposefully exclude truly eligible patients when reporting data to be used in calculating quality measures. Measure design can mitigate the likelihood of exclusions being used inappropriately.

Action to Consider: Develop guidance concerning the use of measure exclusions in collaboration with the National Quality Forum. The guidance would address both new measure development and comprehensive review processes. The guidance would provide clear criteria about provider discretion exclusions, align the use of exclusions in measure specifications for optimal usability, and decrease the likelihood of inappropriate use.

There are various strategies for identifying undesired effects that include data audits, looking at the distributional effects of performance payouts (tied to measures), and conducting qualitative interviews and surveys with providers to ascertain their experiences with the measures. The findings from the systematic review informed the development of qualitative interview guides and provider surveys that CMS plans to conduct as part of the 2018 Impact Report; the purpose of the surveys is to generate national estimates on the type and prevalence of unintended consequences.

Action to Consider: Conduct periodic national provider surveys to develop estimates of the type and prevalence of unintended consequences associated with measure use.

Reference List

- (1) Baker DW, Qaseem A, Reynolds PP, Gardner LA, Schneider EC. Design and use of performance measures to decrease low-value services and achieve cost-conscious care. *Ann Intern Med.* 2013;158(1):55-59.
- (2) Institute of Medicine (US) Roundtable on Evidence-Based Medicine. Yong PL, Saunders RS, Olsen L, ed. *The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary.* Washington, DC: National Academies Press; 2010.
- (3) U.S. Department of Health and Human Services. Totten A, Wagner J, Tiwari A, O'Haire C, Griffin J, Walker M, investigators. *Public Reporting as a Quality Improvement Strategy. Closing the Quality Gap: Revisiting the State of the Science.* Rockville, Md: Agency for Healthcare Research and Quality: Evidence Report No. 208. (Prepared by the Oregon Evidence-based Practice Center under Contract No. 290-2007-10057-I.). *Evid Rep Technol Assess (Full Rep 2012)*; AHRQ Publication No. 12-E011-EF(208.5):1-645.
- (4) Casalino LP. The unintended consequences of measuring quality on the quality of medical care. *N Engl J Med.* 1999;341(15):1147-1150.
- (5) Smith P. On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration.* 1995;18(2/3):277-310.
- (6) Casalino LP, Alexander GC, Jin L, Konetzka RT. General internists' views on pay-for-performance and public reporting of quality scores: a national survey. *Health Aff (Millwood).* 2007;26(2):492-499.
- (7) Ganz DA, Wenger NS, Roth CP, et al. The effect of a quality improvement initiative on the quality of other aspects of health care: the law of unintended consequences? *Med Care.* 2007;45(1):8-18.
- (8) Wachter RM, Flanders SA, Fee C, Pronovost PJ. Public reporting of antibiotic timing in patients with pneumonia: lessons from a flawed performance measure. *Ann Intern Med.* 2008;149(1):29-32.
- (9) Weston A, Caldera K, Doron S. Surgical care improvement project in the value-based purchasing era: more harm than good? *Clin Infect Dis.* 2013;56(3):424-427.
- (10) Damberg CL, Sorbero ME, Lovejoy SL, Martsolf GR, Raaen L, Mandel D. *Measuring Success in Health Care - Value-Based Purchasing Programs: Summary and Recommendations.* Santa Monica, CA: RAND Corporation. RR-306/1-ASPE, 2013.
- (11) Doran T, Kontopantelis E. Pay-for-performance: impact on diabetes. *Curr Diab Rep.* 2013;13(2):196-204.
- (12) Mannion R, Braithwaite J. Unintended consequences of performance measurement in healthcare: 20 salutary lessons from the English National Health Service. *Intern Med J.* 2012;42(5):569-574.

- (13) Werner RM, Asch DA. The unintended consequences of publicly reporting quality information. *JAMA*. 2005;293(10):1239-1244.
- (14) Mannion R, Davies H, Marshall M. Impact of star performance ratings in English acute hospital trusts. *J Health Serv Res Policy*. 2005;10(1):18-24.
- (15) Friedberg MW, Mehrotra A, Linder JA. Reporting hospitals' antibiotic timing in pneumonia: adverse consequences for patients? *Am J Manag Care*. 2009;15(2):137-144.
- (16) Kanwar M, Brar N, Khatib R, Fakhri MG. Misdiagnosis of community-acquired pneumonia and inappropriate utilization of antibiotics: side effects of the 4-h antibiotic administration rule. *Chest*. 2007;131(6):1865-1869.
- (17) Welker JA, Huston M, McCue JD. Antibiotic timing and errors in diagnosing pneumonia. *Arch Intern Med*. 2008;168(4):351-356
- (18) Bloomfield HE, Greer N, Newman D et al. Predictors and Consequences of Severe Hypoglycemia in Adults with Diabetes - A Systematic Review of the Evidence [Internet]. *VA Evidence-based Synthesis Program Reports* 2012.
- (19) Saint S, Meddings JA, Calfee D, Kowalski CP, Krein SL. Catheter-associated urinary tract infection and the Medicare rule changes. *Ann Intern Med*. 2009;150(12):877-884.
- (20) Healy D, Cromwell J. *Hospital-Acquired Conditions-Present on Admission: Examination of Spillover Effects and Unintended Consequences*. Final Report. Research Triangle Park, NC: RTI International; September 2012.
- (21) Campbell S, Hannon K, Lester H. Exception reporting in the Quality and Outcomes Framework: views of practice staff - a qualitative study. *Br J Gen Pract*. 2011;61(585):183-189.
- (22) Doran T, Fullwood C, Reeves D, Gravelle H, Roland M. Exclusion of patients from pay-for-performance targets by English physicians. *N Engl J Med*. 2008;359(3):274-284.
- (23) Casalino LP, Elster A, Eisenberg A, Lewis E, Montgomery J, Ramos D. Will pay-for-performance and quality reporting affect health care disparities? *Health Aff (Millwood)*. 2007;26(3):w405-w414.
- (24) Higgins JP, Green S., eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0. Oxford, UK: The Cochrane Collaboration; updated March 2011.
- (25) von Elm E, Altman D, Egger M, Pocock S, Gøtzsche P, Vandenbroucke J. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol*. 2008;61(4):344-349.
- (26) Methods Reference Guide for Effectiveness and Comparative Effectiveness Reviews. Agency for Healthcare Research and Quality Web site. http://effectivehealthcare.ahrq.gov/repFiles/2007_10DraftMethodsGuide.pdf. Published October 2007. Version 1.0. Accessed September 10, 2014.

- (27) Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med.* 1997;127(8 Pt 2):757-763.
- (28) Rosenthal MB, Frank RG, Li Z, Epstein AM. Early experience with pay-for-performance: from concept to practice. *JAMA.* 2005;294(14):1788-1793.
- (29) McClellan M, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. *JAMA.* 1994;272(11):859-866.
- (30) Owens DK, Lohr KN, Atkins D et al. AHRQ series paper 5: grading the strength of a body of evidence when comparing medical interventions--agency for healthcare research and quality and the effective health-care program. *J Clin Epidemiol.* 2010;63(5):513-523.
- (31) Jha AK, Joynt KE, Orav EJ, Epstein AM. The long-term effect of premier pay for performance on patient outcomes. *N Engl J Med.* 2012;366(17):1606-1615.
- (32) Ryan AM. Effects of the Premier Hospital Quality Incentive Demonstration on Medicare patient mortality and cost. *Health Serv Res.* 2009;44(3):821-842.
- (33) Glickman SW, Ou FS, DeLong ER et al. Pay for performance, quality of care, and outcomes in acute myocardial infarction. *JAMA.* 2007;297(21):2373-2380.
- (34) Centers for Medicare & Medicaid Services, Yale New Haven Health Services Corporation/Center for Outcomes Research and Evaluation (YNHHSC/CORE). *Medicare Hospital Quality Chartbook 2012 Performance Report on Outcome Measures.* September 2012.
- (35) Thompson ND, Yeh LL, Magill SS, Ostroff SM, Fridkin SK. Investigating systematic misclassification of central line-associated bloodstream infection (CLABSI) to secondary bloodstream infection during health care-associated infection reporting. *Am J Med Qual.* 2013;28(1):56-59.
- (36) Reeves MJ, Grau-Sepulveda MV, Fonarow GC, Olson DM, Smith EE, Schwamm LH. Are quality improvements in the get with the guidelines: stroke program related to better care or better data documentation? *Circ Cardiovasc Qual Outcomes.* 2011;4(5):503-511.
- (37) Farmer SA, Black B, Bonow RO. Tension between quality measurement, public quality reporting, and pay for performance. *JAMA.* 2013;309(4):349-350.
- (38) Bernheim SM, Wang Y, Bradley EH et al. Who is missing from the measures? Trends in the proportion and treatment of patients potentially excluded from publicly reported quality measures. *Am Heart J.* 2010;160(5):943-950.
- (39) Ryan AM, Blustein J, Doran T, Michelow MD, Casalino LP. The effect of Phase 2 of the Premier Hospital Quality Incentive Demonstration on incentive payments to hospitals caring for disadvantaged patients. *Health Serv Res.* 2012;47(4):1418-1436.
- (40) Jha AK, Orav EJ, Epstein AM. The effect of financial incentives on hospitals that serve poor patients. *Ann Intern Med.* 2010;153(5):299-306.

- (41) Ryan AM. Has pay-for-performance decreased access for minority patients? *Health Serv Res.* 2010;45(1):6-23.
- (42) Werner RM, Goldman LE, Dudley RA. Comparison of change in quality of care between safety-net and non-safety-net hospitals. *JAMA.* 2008;299(18):2180-2187.
- (43) Campbell SM, Reeves D, Kontopantelis E, Sibbald B, Roland M. Effects of pay for performance on the quality of primary care in England. *N Engl J Med.* 2009;361(4):368-378.
- (44) Campbell SM, Kontopantelis E, Reeves D et al. Changes in patient experiences of primary care during health service reforms in England between 2003 and 2007. *Ann Fam Med.* 2010;8(6):499-506.
- (45) Sutton M, Elder R, Guthrie B, Watt G. Record rewards: the effects of targeted quality incentives on the recording of risk factors by primary care providers. *Health Econ.* 2010;19(1):1-13.
- (46) Mullen KJ, Frank RG, Rosenthal MB. Can you get what you pay for? Pay-for-performance and the quality of healthcare providers. *Rand J Econ.* 2010;41(1):64-91.
- (47) Guthrie B, Auerback G, Bindman AB. Health plan competition for Medicaid enrollees based on performance does not improve quality of care. *Health Aff (Millwood).* 2010;29(8):1507-1516.
- (48) Fagan PJ, Schuster AB, Boyd C et al. Chronic care improvement in primary care: evaluation of an integrated pay-for-performance and practice-based care coordination program among elderly patients with diabetes. *Health Serv Res.* 2010;45(6 Pt 1):1763-1782.
- (49) Doran T, Kontopantelis E, Valderas JM et al. Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK Quality and Outcomes Framework. *BMJ.* 2011;342:d3590.
- (50) Roth S, Gonzales R, Harding-Anderer T et al. Unintended consequences of a quality measure for acute bronchitis. *Am J Manag Care.* 2012;18(6):e217-e224.
- (51) Serumaga B, Ross-Degnan D, Avery AJ et al. Effect of pay for performance on the management and outcomes of hypertension in the United Kingdom: interrupted time series study. *BMJ.* 2011;342:d108.
- (52) Simpson CR, Hannaford PC, McGovern M et al. Are different groups of patients with stroke more likely to be excluded from the new UK general medical services contract? A cross-sectional retrospective analysis of a large primary care population. *BMC Fam Pract.* 2007;8:56.
- (53) Dalton AR, Alshamsan R, Majeed A, Millett C. Exclusion of patients from quality measurement of diabetes care in the UK pay-for-performance programme. *Diabet Med.* 2011;28(5):525-531.

- (54) Sigfrid LA, Turner C, Crook D, Ray S. Using the UK primary care Quality and Outcomes Framework to audit health care equity: preliminary data on diabetes management. *J Public Health (Oxf)*. 2006;28(3):221-225.
- (55) Shah SM, Carey IM, Harris T, Dewilde S, Cook DG. Quality of chronic disease care for older people in care homes and the community in a primary care pay for performance system: retrospective study. *BMJ*. 2011;342:d912.
- (56) Doran T, Fullwood C, Gravelle H et al. Pay-for-performance programs in family practices in the United Kingdom. *N Engl J Med*. 2006;355(4):375-384.
- (57) Kontopantelis E, Doran T, Gravelle H, Goudie R, Siciliani L, Sutton M. Family doctor responses to changes in incentives for influenza immunization under the U.K. Quality and Outcomes Framework pay-for-performance scheme. *Health Serv Res*. 2012;47(3 Pt 1):1117-1136.
- (58) Doran T, Kontopantelis E, Fullwood C, Lester H, Valderas JM, Campbell S. Exempting dissenting patients from pay for performance schemes: retrospective analysis of exception reporting in the UK Quality and Outcomes Framework. *BMJ*. 2012;344:e2405.
- (59) Doran T, Fullwood C, Kontopantelis E, Reeves D. Effect of financial incentives on inequalities in the delivery of primary clinical care in England: analysis of clinical activity indicators for the quality and outcomes framework. *Lancet*. 2008;372(9640):728-736.
- (60) Yu HC, Tsai WC, Kung PT. Does the pay-for-performance programme reduce the emergency department visits for hypoglycaemia in type 2 diabetic patients? *Health Policy Plan*. 2013.
- (61) Petersen LA, Simpson K, Pietz K et al. Effects of individual physician-level and practice-level financial incentives on hypertension care: a randomized trial. *JAMA*. 2013;310(10):1042-1050.
- (62) Kerr EA, Lucatoro MA, Holleman R, Hogan MM, Klamerus ML, Hofer TP. Monitoring performance for blood pressure management among patients with diabetes mellitus: too much of a good thing? *Arch Intern Med*. 2012;172(12):938-945.
- (63) Beard AJ, Hofer TP, Downs JR et al. Assessing appropriateness of lipid management among patients with diabetes mellitus: moving from target to treatment. *Circ Cardiovasc Qual Outcomes*. 2013;6(1):66-74.
- (64) Lai CL, Hou YH. The association of clinical guideline adherence and pay-for-performance among patients with diabetes. *J Chin Med Assoc*. 2013;76(2):102-107.
- (65) Chien AT, Eastman D, Li Z, Rosenthal MB. Impact of a pay for performance program to improve diabetes care in the safety net. *Prev Med*. 2012;55 Suppl:S80-S85.
- (66) Chang RE, Lin SP, Aron DC. A pay-for-performance program in Taiwan improved care for some diabetes patients, but doctors may have excluded sicker ones. *Health Aff (Millwood)*. 2012;31(1):93-102.

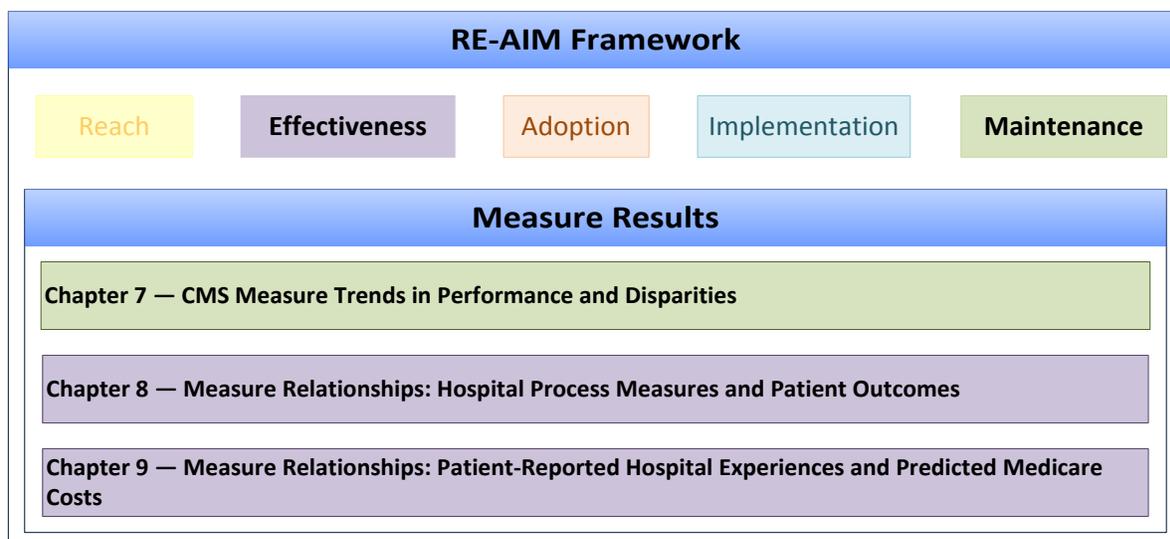
- (67) Millett C, Netuveli G, Saxena S, Majeed A. Impact of pay for performance on ethnic disparities in intermediate outcomes for diabetes: a longitudinal study. *Diabetes Care*. 2009;32(3):404-409.
- (68) Lee JT, Netuveli G, Majeed A, Millett C. The effects of pay for performance on disparities in stroke, hypertension, and coronary heart disease management: interrupted time series study. *PLoS One*. 2011;6(12):e27236.
- (69) Hamilton FL, Bottle A, Vamos EP et al. Impact of a pay-for-performance incentive scheme on age, sex, and socioeconomic disparities in diabetes management in UK primary care. *J Ambul Care Manage*. 2010;33(4):336-349.
- (70) Alshamsan R, Lee JT, Majeed A, Netuveli G, Millett C. Effect of a UK pay-for-performance program on ethnic disparities in diabetes outcomes: interrupted time series analysis. *Ann Fam Med*. 2012;10(3):228-234.
- (71) Konetzka RT, Polsky D, Werner RM. Shipping out instead of shaping up: rehospitalization from nursing homes as an unintended effect of public reporting. *J Health Econ*. 2013;32(2):341-352.
- (72) Werner RM, Konetzka RT, Stuart EA, Polsky D. Changes in patient sorting to nursing homes under public reporting: improved patient matching or provider gaming? *Health Serv Res*. 2011;46(2):555-571.
- (73) Werner RM, Konetzka RT, Stuart EA, Norton EC, Polsky D, Park J. Impact of public reporting on quality of postacute care. *Health Serv Res*. 2009;44(4):1169-1187.
- (74) Cooke V, Arling G, Lewis T, Abrahamson KA, Mueller C, Edstrom L. Minnesota's Nursing Facility Performance-Based Incentive Payment Program: an innovative model for promoting care quality. *Gerontologist*. 2010;50(4):556-563.

Part 2—Overview of CMS Quality Measure Results: Effectiveness and Maintenance

Part 2 of the report provides detailed analyses of quality measure results over time within and across the Centers for Medicare & Medicaid Services (CMS) measure programs. The analyses place special emphases on identifying disparities in quality of care among specific populations of interest. Part 2 also includes analyses that examine the relationships among measures within the CMS programs. The aim of this section is to identify patterns, gaps, characteristics, insights, and lessons from measure results to help calibrate the development and use of quality measures in the investigated programs, and, by extension, in other CMS programs.

The topics for the chapters are presented in Figure P-2, which also indicates the corresponding aspect of the RE-AIM Framework. Each chapter begins with the research question(s) and includes the programs and care settings addressed by the chapter. As in Part 1, each chapter also includes Background, Methods, Results, and Conclusions.

Figure P-2: Chapter Titles Based on Research Questions



Chapter 7—CMS Measure Trends in Performance and Disparities



Questions on Maintenance

1. To what extent did providers' performance on the quality measures improve over time?
2. What are the disparities in measure rates for age, sex, race, and ethnicity, and how did these disparities change over time?

Abstract

Background: The Centers for Medicare & Medicaid Services (CMS) collects and reports hundreds of measures to assess the quality of care across hospital, ambulatory, and post-acute care settings as well as across CMS measurement programs. These include the Hospital Inpatient Quality Reporting Program, the Hospital Outpatient Quality Reporting Program, the Nursing Home Quality Initiative, the Home Health Quality Reporting Program, Medicare Part C, and the End Stage Renal Disease Quality Incentive Program. This research quantifies the improvement in quality of care as represented by CMS quality measures, details the differences in performance on quality measures by provider characteristics, describes disparities in measure rates between patient demographic groups, and tests for changes in identified disparities over time.

Methods: Using data collected across providers and patients, the research team examined trends in the provider performance rates of 119 quality measures for which three or more years of data were available across seven CMS programs between 2006 and 2012. The research team implemented a decision tree to classify measures as high performing (with rates of > 90 percent when higher rates are desirable or < 5 percent when lower rates are desirable over three consecutive measurement years) or not high performing and as exhibiting one of four levels of change (from substantial increases to substantial decreases in rates). To measure changes over time, the research team used Cohen's *d* as a measure of effect size, defining $d > 0$ as "improvement" and $d \geq 0.2$ as a "substantial improvement." Differences in provider performance on the quality measures across provider

characteristics and patient demographic groups were described using these same metrics. The research team performed a disparity analysis using a methodology based on that of the *National Healthcare Disparities Report* to determine if the differences in quality measure rates across patient demographic groups were diminishing over time. Results for age, sex, and race were reported for 104 measures, while results were reported for 59 measures for ethnicity.

Results: Across 119 quality measures examined, 95 percent showed improvement over a period of three to six years. Approximately 35 percent of measures were classified as high performing. Of those measures that were not high performing, 91 percent exhibited an increase. In each of the seven programs examined, 89 to 100 percent of the measures demonstrated an increase over the three to six years of available data. Thus, broad evidence supports that quality of care and patient outcomes are improving over time, although more process measures than outcome measures were substantially improving over the study period. Seventy-five percent of process measures were classified as high performing or substantially improving compared with 20 percent of outcome measures classified in these same categories.

Differences in provider characteristics and patient demographics associated with measures defined as high performing were observed across all programs. Disparities by sex were found in just 9 percent of measures, by ethnicity in 22 percent, and by age in 37 percent, compared with 48 percent by race and race/ethnicity. Sixty-six percent of the race and race/ethnicity disparities diminished over time, compared with 22 percent of disparities by sex, 42 percent by age, and 77 percent for ethnicity alone. All programs included significant disparities in measure rates, particularly across race and age groups; however, the results also showed that the magnitude of these disparities is diminishing in all programs except for the Home Health Quality Reporting Program.

Conclusions: The results suggest that the CMS quality measurement programs are associated with improvements in care across all programs and demographic groups. More process measures improved during than study period than outcome measures, and 75 percent of process measures were either classified as high performing or substantially improving. While CMS is phasing out process measures and moving toward developing additional outcome measures, retaining process measures will provide timely information on the progress of quality improvement efforts. Although disparities in measure rates are diminishing, they continue to persist across programs, settings, and demographic groups. Strategies to eliminate disparities should include explicit monitoring of measure rates of racial and ethnic groups as well as uniform collection of racial/ethnicity data across all measurement programs.

Background

For over a decade, the Centers for Medicare & Medicaid Services (CMS) has been collecting and reporting standardized measures of the quality of healthcare services as a strategy for increasing transparency and promoting improvements in care delivery. From collecting data on a small number of measures used to evaluate performance in managed care plans in the 1990s to collecting data on hundreds of measures today, CMS is assessing the quality of care delivered in a range of programs and settings, including those for hospitals, nursing homes, home health agencies, Medicare Part C plans, and dialysis facilities. This study quantitatively assesses provider performance between 2006 and 2012 in the aspects of healthcare and outcomes tracked by CMS quality measures.

Objectives

Four objectives guided the analysis to examine performance trends and disparities over time:

1. Analyze the absolute improvement of measure rates.
2. Classify nationally aggregated measure rates by levels of performance and improvement.
3. Analyze the performance and improvement of measure rates by provider characteristics: Affiliation, ownership, plan type, safety net status, staffing hours, teaching status, and the degree to which an area is urban (urbanicity).
4. Quantitatively assess disparities in initial measure rates by age, sex, race and ethnicity and determine whether disparities diminish over the study period.

Methods

Programs, Measures, and Data Sources

To address Objectives 1 through 4, the research team assessed the CMS quality measure results in terms of levels of providers' performance and the extent of improvement of the measure rates over time. The analysis included measures for which three or more consecutive years of data were available for the interval 2006 through 2012. One hundred nineteen measures from seven CMS measurement programs met the criteria for inclusion. A hyperlink to a complete list of measures used in this analysis is provided in Appendix i-4. The list includes the NQF endorsement status, NQF number if endorsed, and both the measure title used by the CMS program and the measure title used by NQF. The included measures allow the examination of aspects of care in hospital, ambulatory, and post-acute care settings (Table 7-1).

Trends over time for each measure were evaluated, beginning either in 2006 or in the first full year for which data were publicly available. Measures were summarized at the annual level to avoid the complication of seasonal fluctuations. Seasonality is, by definition, systematic. The systematic nature of seasonality serves to mask broader trends, by adding seasonal fluctuations to the rates. If seasonality is not removed from the data, then apparent changes in performance would be partially driven by factors related to seasonality, and the result would potentially be a false representation of performance. Since the goal of this research question is to assess the trends in quality measure rates between 2006 and 2012, annual rates were used to remove the seasonal effect to determine the overarching trend. The trend analysis for each measure ended in

the last complete year for which data were available at the time of analysis. For measures in the Home Health Quality Reporting Program (HH QRP) and the Nursing Home Quality Initiative (NHQI), the quality assessment instruments were replaced in 2009 and 2010, respectively. The instrument changes resulted in significant differences in the specifications of the measures in these two programs, which precluded continuous trend analyses; thus, the research team used data for these two programs from 2006 through 2009. Data for most measures in the remaining programs were publicly available through 2012.

Table 7-1: Programs, Measures, Time Periods, and Data Sources

Setting	Program/ (Abbreviation)	Measures (n)	Time Period	Data Source
 Hospital	Hospital Inpatient Quality Reporting Program (Hospital IQR Program) ^{xi,xli,xlii}	AMI, HF, PN, SCIP Quality Measures (28)	2006–2012	CMS Abstraction & Reporting Tool (CART)
		HCAHPS (10)	2008–2012	Hospital Compare, CMS contractor
		Structural Measures (3)	2009–2012	Hospital Compare
		Mortality and Readmission Measures (6)	2009–2012	CMS contractor
	Hospital Outpatient Quality Reporting Program (Hospital OQR Program) ^{xliii}	Hospital OQR Program Quality Measures (7)	2010–2012	Hospital Compare, CART
 Ambulatory	Medicare Part C (Part C)	MA CAHPS (7)	2007–2013	CMS contractor
		Health Outcomes Survey (6)	2007–2013	CMS contractor
		Healthcare Effectiveness Data and Information Set (HEDIS) (14)	2007–2013	CMS contractor
	Medicare Part D (Part D) ^{xliiv}	Medication Measures (5)	2008–2012	CMS contractor
		PDP CAHPS (3)	2007–2013	CMS contractor
 Post-Acute	Nursing Home Quality Initiative (NHQI)	MDS 2.0 Quality Measures (19)	2006–2009	Medicare.gov Nursing Home Compare, Quality Improvement and Evaluation System (QIES): MDS 2.0
	Home Health Quality Reporting Program (HH QRP)	Outcome and Assessment Information Set (OASIS) B Quality Measures (9)	2006–2009	Home Health Compare, QIES: OASIS-B1 Outcome-Based Quality Improvement (QBQI)
	End-Stage Renal Disease Quality Incentive Program (ESRD QIP)	ESRD Quality Measures (2)	2006–2011	CMS contractor

^{xi} Individual measures within a set (e.g., the Hospital Inpatient Quality Reporting (IQR) Program Acute Myocardial Infarction (AMI), Heart Failure (HF), Pneumonia (PN), and Surgical Care Improvement Project (SCIP) Quality Measures) may not be available for the full range of years reported in the table. Some measures were introduced after 2006, and others were removed prior to 2012. The time periods noted represent the full range of years for all included program measures.

^{xli} When appropriate, the Medicare.gov Hospital, Nursing Home, and Home Health Compare websites were used as data sources. Data were extracted from these sites in December 2012. The Compare files are subject to periodic revision, so it may not be possible to obtain identical results through analysis of currently available data from the Compare websites.

^{xlii} Hospital IQR Program measure time periods refer to the time period of data collection.

^{xliii} Hospital OQR Program measure time periods refer to the time period of data collection.

^{xliiv} All measures other than those associated with Medicare Parts C and D are identified by the year in which the data were collected. By convention, however, Part C and D are identified by the year in which they were reported, which is normally the year after they were collected. Thus, the 2013 Parts C and D data were actually collected in 2012.

Analysis

Objectives 1 and 2—Trend Analysis

Objective 1: Analyze the absolute improvement of measure rates.

Objective 2: Classify nationally aggregated measure rates by levels of performance and improvement.

The 2012 *National Impact Assessment of Medicare Quality Measures Report* (2012 Impact Report) described changes in 72 measures from 2006 through 2010. Two years of data and 47 measures were added to the 2015 *National Impact Assessment of the Centers for Medicare & Medicaid Services (CMS) Quality Measures Report* (2015 Impact Report). The dimensions of performance and improvement were also added. Two primary purposes of quality measurement are to allow assessment of the quality of care providers are delivering to patients at a particular point in time and to assess the degree to which providers are increasing the quality of care delivery over time. The first dimension is *performance*, and the second is *improvement*.

Performance

In this chapter, references to high-performing measures represent the aggregate results of the providers' performance.

Measures were defined as *high performing* if they approached their maximum possible values and had limited room for improvement. Ideally, the performance of a provider on a CMS quality measure would be compared to a benchmark developed specifically for that measure. However, many of the measures examined have no official benchmarks or target ranges. The measures are expressed in rates that vary between 0 percent and 100 percent. Generally, a higher rate is desirable, e.g., for *AMI-1: Aspirin at Arrival for Acute Myocardial Infarction*,^{xlv} a higher percentage of patients receiving aspirin at arrival for acute myocardial infarction is desirable. However, lower rates are more desirable in measures such as *High-Risk Residents with Pressure Ulcers* (NQF #0679) in which a lower percentage of patients with pressure ulcers is desirable.

Because systematic benchmarks or a systematic benchmarking process did not exist for each measure, the research team created a system to define “high performing.” The creation of the benchmark ensured that all measures would be compared against the same standard. To establish this benchmark, the research team met with consultants and agreed that when the desired goal for a measure was 100 percent, a provider that had a score of at least 90 percent during each of the most recent three consecutive years was performing very well. However, the consensus regarding performance on measures in which a lower rate was desirable was that a rate of 10 percent (the converse of 90 percent) was unacceptably high. For example, the research team felt that a rate of 10 percent for *High-Risk Residents with Pressure Ulcers* or *Residents Who Were Physically Restrained (Long Stay)* (NQF #0687) would not qualify as “high performing” for this study. For these reasons, the thresholds of 90 percent for positive measures and 5 percent for negative measures were established.

^{xlv} Previously NQF-endorsed title: *Aspirin at arrival for acute myocardial infarction (AMI)* (NQF #0132).

Improvement

A measure rate improved if the final rate was closer to the desired rate compared to its baseline value. Therefore, “improved” means higher compared to the baseline measurement for a measure where a higher rate is desirable or lower compared to the baseline measurement when a lower rate is desirable. The rate for the Medicare Part C measure *Plan Members With Diabetes Whose Blood Sugar Is Under Control*,^{xlvi} for example, increased from 46 percent in 2007 to 69 percent in 2013, leading to its classification as improved. In contrast, the rate for the NHQI measure *Residents With Pressure Ulcers (Short Stay)*^{xlvii} declined from 17 percent in 2006 to 13 percent in 2009, which is a change in the desired direction leading to its classification as improved.

Substantial Improvement

Statistical significance testing does not address the magnitude of changes in measure rates. To address this limitation, the research team used an effect size metric to define a “substantial” improvement as measure rates that improved to a clinically meaningful degree. All changes in the desired direction are improvements, but not all observed improvements are statistically significant; examples include changes that are the result of minor short-term fluctuations or random errors in the measurement processes.

Tests of statistical significance are used to assess whether an observed improvement is likely to reflect a real change in the measure; however, there are situations where significance testing has limited relevance. In particular, when the number of observations for a measure is large, as is the case with national quality reporting data, small changes can attain statistical significance. These changes may be real in a statistical sense, but they may also be too small to be clinically meaningful. A statistically significant increase in a measure rate may not have an effect size that is large enough to be relevant to quality improvement.

Analysis of change over time will nearly always result in statistical significance for even trivially small differences when datasets are large. For example, if a measure with 100,000 entries (e.g., beneficiaries, hospital admissions) at each of two points in time changed from a rate of 80.00 percent to 80.35 percent, the change would be considered statistically significant at the $p < .05$ level. Yet, given that result, 350 entries out of 100,000 would have been different at time two from what they were at time one. This relative improvement of just 0.4 percent would be highlighted as being important by labeling it as statistically significant.

Even a few data points could show statistical significance for a small change due to nearly linear trends over time. A measure with four data points (e.g., 70.10 percent, 70.14 percent, 70.16 percent, and 70.20 percent) that changed by one-tenth of a percentage point can generate a p value less than .01 for this trivially small change. This time, however, the small p value was generated from the lack of variation (i.e., +.04, +.02, +.04) rather than from large sample sizes. Both issues (i.e., number of units being assessed and the variation, or lack thereof) needed to be accommodated for a policy-relevant impact assessment for the tracked quality measures.

^{xlvi} NQF-endorsed title: *Comprehensive Diabetes Care: Hemoglobin A1c (HbA1c) Poor Control (>9.0%)* (NQF #0059).

^{xlvii} NQF-endorsed title: *Percent of Residents or Patients with Pressure Ulcers That Are New or Worsened (Short-Stay)* (NQF #0678).

Stakeholders (such as patients, clinicians, and insurers) might not agree, however, on how large a change would be policy relevant for the CMS measures. Furthermore, the observed improvement in a rate for a measure—the absolute change between baseline and final measurement—is insufficient to assess the magnitude of an observed change because it is not adjusted for variability. The baseline and final measurement are averages of measure performance rates reported for large numbers of providers or patients. Higher variation in rates for a measure decreases the certainty with which the magnitude of improvement can be established. The statistical concept of “effect size” addresses the issue of determining whether an improvement is large enough to be policy relevant and accounts for variability in the data.¹

A frequently used measure of effect size is Cohen’s *d*. Cohen’s *d* is calculated by dividing the linearized estimate of absolute change over time by a standard measure of variability. The smaller the variability relative to the magnitude of the change over time, the larger Cohen’s *d* grows; a larger Cohen’s *d* indicates a larger effect. While there is no absolute answer to the question of how large Cohen’s *d* needs to be for an effect to be considered substantial or important, Cohen suggested that a value ≥ 0.2 indicates a notable effect. Therefore, this standard is applied to distinguish between measures that show a *substantial* increase or decrease (Cohen’s $d \geq 0.2$) and those that have changed *slightly* (Cohen’s $d < 0.2$).

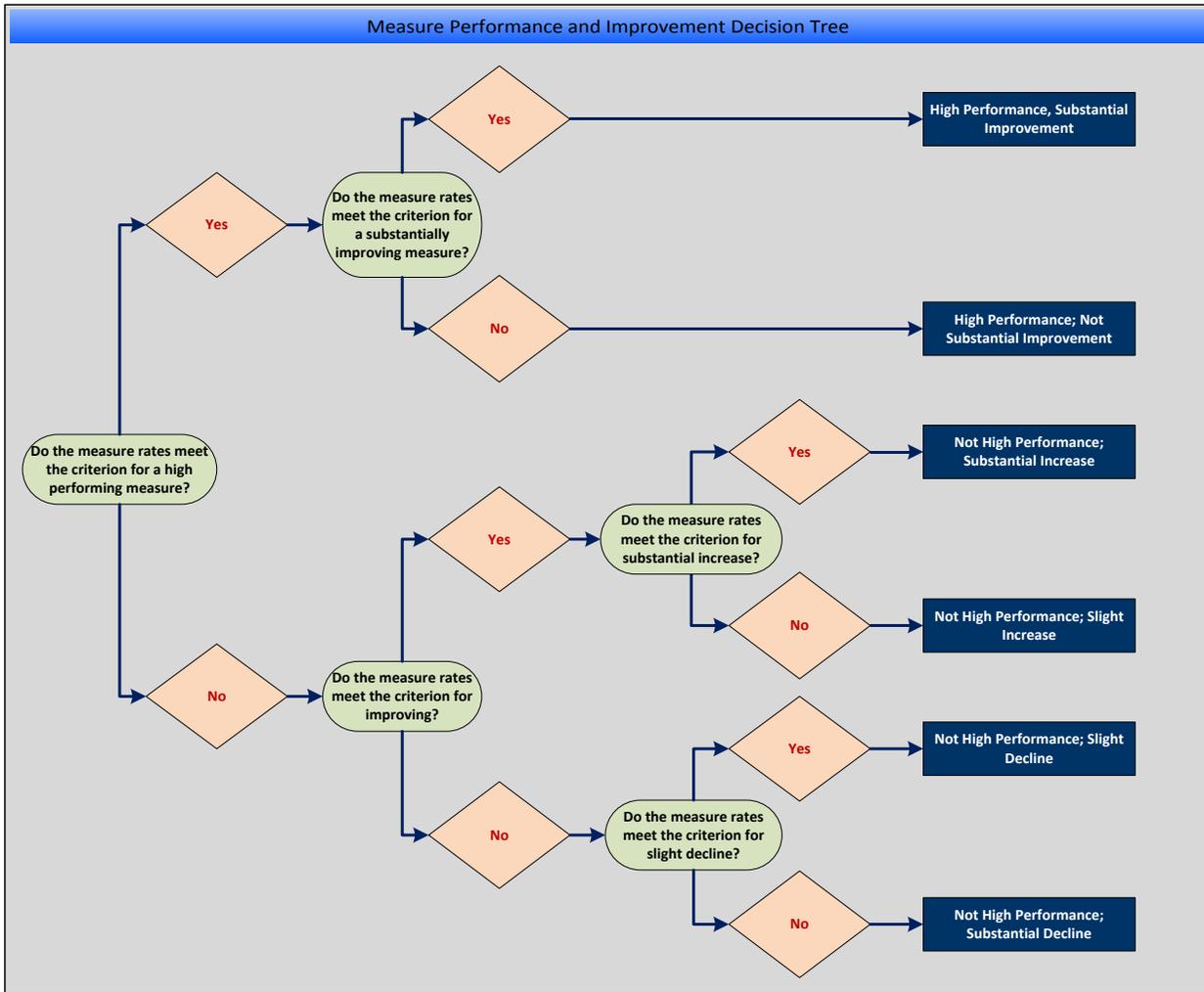
Cohen’s *d* is calculated when an estimate of variability is available. Approximately 50 percent of the CMS quality measures have been reported in such a way that an estimate of variability cannot be calculated (see Appendix 7-1 for further details). Additionally, Cohen’s *d* cannot be calculated for categorical or dichotomously scored variables. When it was not possible to calculate Cohen’s *d*, Annual Percentage Change (APC) was used. Calculating APC for a measure requires fitting a straight line to the observed annual values of the measure using ordinary linear regression. The APC is the slope of the linearized measure trend. The research team determined empirically (see Appendix 7-1 for details) that an APC of 1.4 percent (0.014) for CMS quality measures was equivalent to a Cohen’s *d* of 0.2, so an APC of 1.4 percent was chosen as the threshold for categorizing a measure as substantially improving or deteriorating. Cohen’s *d* was used as the effect size metric in 49.7 percent of the total number of calculations in the analyses of substantial change; APC was used for the other 50.3 percent (Appendix 7-1).

For both Cohen’s *d* and APC, change is computed based on the linearized estimates of the observed rates. This procedure removes random year-to-year fluctuations in the measure rates resulting in a smoother trajectory of change. The smoothed trajectory allows for calculation of a generalizable effect size which is an average across the time period being studied. If the random fluctuations are not smoothed out, then calculations of changes in rates (i.e., trends) will be biased by random noise in the measure.

Combining Performance and Improvement

Each of the CMS quality measures was analyzed and sorted into one of six comprehensive and mutually exclusive categories based on its performance and improvement over the study period. As shown in Figure 7-1, measures meeting the high-performing standards were subdivided into those showing substantial improvement (Cohen’s $d \geq 0.2$) and those not showing substantial improvement (Cohen’s $d < 0.2$). The research team divided the remaining measures according to whether they were improving or declining and then according to whether they were improving or declining substantially or slightly using the same criteria for Cohen’s *d*.

Figure 7-1: Performance and Improvement Assignment Decision Tree



Reporting on Performance and Improvement

Categorized measures were tabulated by care setting (hospital, ambulatory, post-acute), program (e.g., Hospital Inpatient Quality Reporting (IQR) Program, NHQI), measure type (process, outcome, structure), provider characteristics (e.g., size, geographic location), and patient characteristics (e.g., sex, age). For the reasons described in the methods section defining the use of Cohen’s d for substantial improvement, no tests of statistical significance were conducted on differences across categories. That is, the percentage of Hospital IQR Program measures that are improving might be higher than the percentage of NHQI measures that are improving, but whether the observed differences between these programs are meaningful cannot necessarily be inferred.

Trends are displayed graphically for a small selection of measures to illustrate major themes revealed by the analysis and to serve as a reminder of the individual measures that underlie the main tables.

Objective 3—Provider Characteristics

Objective 3: Analyze the performance and improvement of measure rates by provider characteristics.

The primary sources of the provider characteristics were the respective Compare websites. Other sources included the American Hospital Association Survey, the CMS Health Care Quality Improvement System (QIES), CMS Abstraction Reporting Tool (CART), and CMS subcontractors. The provider characteristics (e.g., nursing home affiliation) were then either used as identified through the original data sources or further calculation was performed as necessary (e.g., urbanicity of hospitals). Table 7-2 describes further calculations, if they were needed.

Table 7-2: Data Sources and Methods Used in Provider Characteristics Analyses

Characteristic ^{xlvi}	Source	Note
Hospital Characteristics		
Urbanicity	Hospital Compare	Cross-referenced ZIP code of facility to 2010 Census to designate facilities as urban or rural
Ownership	Hospital Compare	Categorized by Hospital Compare
Teaching Status	American Hospital Association Survey (AHA Survey)	Categorized by AHA Survey
Safety Net	AHA Survey	Safety Net hospitals were identified as hospitals that had Medicaid inpatient utilization rate one standard deviation or more above the mean for all hospitals in the state ²
Bed Size	AHA Survey	Hospitals were grouped according to bed sizes identified in the AHA Survey
Nursing Home Characteristics		
Ownership	Nursing Home Compare (NH Compare)	Categorized by NH Compare
Affiliation	NH Compare	Categorized by NH Compare
Bed Size	NH Compare/CMS Health Care Quality Improvement System (QIES)	Size calculated in quartiles of the number of beds identified in NH Compare and/or QIES
Nursing Hours	NH Compare	Nursing hours were calculated in quartiles based on the number of nursing hours per resident identified in NH Compare
Urbanicity	NH Compare/QIES	Cross-walked ZIP code of facility to 2010 Census to designate facilities as urban or rural

^{xlvi} All characteristics were identified during the baseline period for each facility, unless otherwise indicated. The baseline period would be the first year the facility appeared in the data files.

Table 7-2: Data Sources and Methods Used in Provider Characteristics Analyses

Characteristic ^{xlviii}	Source	Note
Home Health Characteristics		
Ownership	Home Health Compare/QIES	Categorized by Home Health Compare and/or QIES
Size	QIES	Size calculated in quartiles of episodes of care identified in QIES
Urbanicity	QIES	Since a home health agency (HHA) does not necessarily care for patients at its office location, QIES was used to identify the urbanicity of the clients. If a HHA had greater than 50 percent of <i>clients</i> in the baseline year that were urban, then the HHA was designated as urban.
Part C Characteristics		
Ownership	CMS Subcontractor	
Urbanicity	CMS Subcontractor	
Plan Type	CMS Subcontractor	
SNF Type	CMS Subcontractor	
Part D Characteristics		
No provider characteristics presented in Chapter 7		

Objective 4—Disparities Analysis

Objective 4: Quantitatively assess disparities in initial measure rates by age, sex, race, and ethnicity and determine whether disparities diminish over the study period.

Healthcare disparities exist when disadvantaged or vulnerable groups receive lower quality services or experience poorer outcomes relative to comparatively advantaged reference groups. Eliminating healthcare disparities by age, sex, and race and ethnicity is an important foundational principle of the CMS Quality Strategy.³ This part of the research was designed to answer three questions. The three questions are:

1. Were there differences in performance and improvement on quality measure rates for the demographic groups defined by age, sex, and race and ethnicity?
2. Did large differences exist in the quality measure results between the demographic groups?
3. Where disadvantaged or vulnerable groups experienced disparities, was the gap between these groups and the reference groups increasing or diminishing over time?

The research team addressed the first question by applying the same methodology as that used to classify performance and improvement on the quality measure rates for age, sex, and race and ethnicity. The second and third questions regarding disparities in care and the trajectories of those disparities over time were addressed using methods based on the *2013 Agency for Healthcare Research and Quality (AHRQ) National Healthcare Disparities Report (NHDR)*.⁴ Disparities were ascertained by comparing performance on the quality measure for a reference group to its performance for a comparison group. A disparity existed between a reference group and a comparison group when the baseline measurements for the two groups differed significantly ($p < 0.05$, two-tailed) and the difference was 5 percentage points or more in an

unfavorable direction compared to the reference group. For example, on the measure *Surgery Patients Who Received Appropriate Venous Thromboembolism Prophylaxis Within 24 Hours Prior to Surgery to 24 Hours After Surgery* (NQF #0218), 66 percent of Hispanic patients received the standard of care compared with 78 percent of non-Hispanics. This is a statistically significant unfavorable difference of 5 percentage points or greater because fewer Hispanics received the standard of care.

A disparity between a comparison and a reference group is considered to be diminishing over time if the average annual rate of change is greater than 1 percent in the desirable direction and the difference in slopes is significant at the $p < 0.05$ level. For *Surgery Patients Who Received Appropriate Venous Thromboembolism Prophylaxis Within 24 Hours Prior to Surgery to 24 Hours After Surgery*, by 2012, 96 percent of Hispanics and 98 percent of non-Hispanics received the standard of care. The measure rate for Hispanics improved 30 percentage points compared with the 20 percent for non-Hispanics, indicating that the disparity has diminished.

AHRQ used slightly different rules for identifying disparities and defining diminishing disparities over time. AHRQ defined a 10-percentage-point difference between reference and comparison groups rather than a 5-percentage-point difference. When determining if a disparity has diminished, AHRQ used a less conservative threshold of $p < 0.10$, while this study set $p < 0.05$ for testing the significance of the difference between reference and comparison group measure trend slopes. As a result, the AHRQ method was less likely to identify disparities and more likely to consider existing disparities to be diminishing over time.

AHRQ also excluded measures that had a measure rate greater than 95 percent. In this study, such measures are classified as high performing and are considered an important part of the story of progress in meeting quality objectives. Omitting these may have led AHRQ to report less progress in diminishing disparities. In addition, the sole focus of this research is CMS quality measures. While the AHRQ NHDR included CMS quality measures, a number of other measures were used, including clinical conditions measures, measures of maternal and child health, and measures of workplace diversity.

For the disparities analysis, the age groups were 18–64, 65–84, and 85+, with 65–84 as the reference group. The sex categories were male and female, with male as the reference group. The race and ethnicity categories varied by program (see discussion under “Race and Ethnicity”): White was the reference category for race, non-Hispanic White was the reference category for race/ethnicity, and non-Hispanic was the reference category for ethnicity.

The research team evaluated each measure for each pair of comparison and reference groups. For example, for the ambulatory care process measure *Colorectal Cancer Screening* (NQF #0034), the 85+ age group was compared to the 65–84 age group. If a disparity was detected in one or both of these comparisons, an age disparity would be reported for this measure, but the number of disparities found by age would not be reported. For changes over time in disparities, if there were disparities for both the younger (18–64) and older (85+) age groups and either disparity diminished, it would be reported that a disparity had diminished. While it is hypothetically possible for one disparity to diminish, and another to increase, this scenario was never observed in the data. Therefore, a diminishing disparity represents the reduction of differences between one or both demographic groups and the reference group.

Race and Ethnicity

Standards for the collection of information about patients’ race and ethnicity have been developed (see 1997 standards published by OMB⁵ and 2011 standards put forth by HHS in response to ACA section 4302⁶); however, the challenge is that these standards are not consistently adopted across measurement programs. CMS quality measurement programs collect and organize data on the race and Hispanic ethnicity of patients (Table 7-3) differently. Programs such as Medicare Part C (Part C), Medicare Part D (Part D), NHQI, and HH QRP collect data on race and ethnicity as a single data element without making a distinction between the two. When race and ethnicity are treated as a single combined category, as they are in all programs except the Hospital IQR and the Hospital Outpatient Quality Reporting (OQR), that category is referred to as “race/ethnicity.” An ethnicity variable was constructed for NHQI and HH QRP, based on the data available in the combined race/ethnicity variable. The Hospital IQR Program and Hospital OQR Program collect race and ethnicity as distinct data elements, so the categories are referred to separately for these programs. While the HCAHPS survey is part of these hospital programs, HCAHPS collects information on patients’ race and ethnicity differently than the Hospital IQR Program and Hospital OQR Program, as presented in Table 7-3.

Table 7-3: Race or Race/Ethnicity Variables Present in the Program Data or Constructed for Analysis

Program/Source	Race or Race/Ethnicity	Ethnicity
Hospital IQR Program, Hospital OQR Program	<ol style="list-style-type: none"> American Indian or Alaska Native (AIAN) Asian (Asian) Black (Black) Native Hawaiian or Pacific Islander (NHPI) White (White) 	<ol style="list-style-type: none"> Hispanic Non-Hispanic
Part C, Part D, HCAHPS ^{xlix}	<ol style="list-style-type: none"> American Indian or Alaska Native (AIAN) Asian/Pacific Islander (Asian/PI) Black (Black) Hispanic (Hispanic) Non-Hispanic White (White) Multicultural/Other/Unknown (Other) 	
NHQI	<ol style="list-style-type: none"> American Indian/Alaska Native (AIAN) Asian/Pacific Islander (Asian/PI) Black, not of Hispanic origin (Black) Hispanic (Hispanic) White, not of Hispanic origin (White) Unknown (Other) 	<ol style="list-style-type: none"> Hispanic* Non-Hispanic*
HH QRP	<ol style="list-style-type: none"> American Indian or Alaska Native (AIAN) Asian (Asian) Black or African American (Black) Hispanic (Hispanic) Native Hawaiian or Pacific Islander (NHPI) White (White) Unknown Race/Ethnicity (Other) 	<ol style="list-style-type: none"> Hispanic* Non-Hispanic*

*These are constructed variables.

^{xlix} Patient-level data were not available. A separate ethnicity category could not be created.

Limitations

This chapter focuses on CMS quality measures and does not provide a fully representative overview of the quality of care delivered to the U.S. population across all age groups, conditions, or care settings. The analyses were limited by the particular characteristics of the data that were available for CMS measurement programs, and it is beyond the scope of this research to examine all aspects of measure performance. A measure is said to be “high performing” if the measure results meet the criteria established in this study. This research does not address the completeness or accuracy of data reported to CMS for a measure, nor does it address other aspects of measure performance not mentioned here.

As discussed in Chapter 1, hospital quality measures from the Hospital IQR Program and the Hospital OQR Program are greatly outnumbered by ambulatory care measures from the Physician Feedback Program and the Physician Quality Reporting System (PQRS). Due to data availability, however, 45 percent of the 119 measures included in the trend analyses in this 2015 Impact Report are from the Hospital IQR Program and the Hospital OQR Program, and 29 percent are ambulatory measures. This weights the results toward hospitals, giving a more complete picture of hospital care than other types of care. Likewise, the focus on CMS measures means that greater insight is provided into the quality of care received by Medicare patients than that received by the general public.

Suitable data were not available for all CMS quality measurement programs for the study period, primarily because newer programs have yet to generate enough consecutive years of data to meet inclusion criteria. This means that no quality measures were included for these programs, e.g., hospice care; therefore, the results cannot be comprehensive across all CMS programs.

The diversity of ways in which programs report the racial and ethnic identity of patients posed challenges and complicated the interpretation of results. Caution should be applied in making generalizations about disparities across settings for several reasons. First, programs differ in how data on race and ethnicity are collected, so one-to-one program correspondence cannot be achieved. Second, the manner in which this information is collected varies across settings, which introduces an element of unreliability. The extent of this type of unreliability was not investigated for this study. A third reason to be cautious about making generalizations regarding disparities across settings is that each setting has different data-entry processes that may not be equally reliable.

Measure specifications change over time (for an illustration, see Chapter 8, Figure 8-3: *HF-1: Discharge Instructions¹ Denominator Population Versus Measure Specifications in 2006–2012*). Although certain changes are substantial (e.g., potentially doubling the number of patients whose care or outcomes are being measured), the majority of specification changes are limited in scope and do not result in substantial changes in the population being measured. The statistical methods used for this trend analysis rely on the assumption that the measure is the same on a year-to-year basis. Therefore, this limitation needs to be considered when interpreting the findings of the trend analysis.

¹ Previously NQF-endorsed title: *Heart Failure (HF): Detailed discharge instructions* (NQF #0136).

Finally, while a greater quantity of quality measures were included in the 2015 Impact Report than in the 2012 Impact Report, many measures of interest were not included because either data were publicly unavailable or the measures were too new to meet the three-year minimum standard for calculating trends. The 2018 Impact Report will incorporate additional measures as they become available, facilitating understanding of a broader range of CMS quality measures.

Results

Objectives 1 and 2—Overall Performance and Improvement Trends

Objective 1: Analyze the absolute improvement of measure rates.

Objective 2: Classify nationally aggregated measure rates by levels of performance and improvement.

Table 7-4 summarizes the number of measures included by program and by measure type along with the percentage of those measures that demonstrated improvement over the study period. A measure is considered to have improved if the final measurement is better than the baseline measurement. In the period under review (2006–2012), 95 percent of the measures improved. The list of included measures is in Appendix 7-2 through Appendix 7-21.

Table 7-4: Number of Quality Measures With Improved Rates by Program and Measure Type

Program	N	Percent Improved
Hospital: Hospital IQR Program	47	95.7%
Hospital: Hospital OQR Program ⁱⁱ	7	100%
Ambulatory: Part C	27	96.3%
Ambulatory: Part D	8	100%
Post-Acute: HH QRP	9	88.9%
Post-Acute: NHQI	19	89.5%
Post-Acute: ESRD QIP	2	100%
Total	119	95.0%
Type	N	Percent Improved
Outcome: Clinical	25	88.0%
Outcome: Access	1	100%
Outcome: Efficiency ⁱⁱⁱ	0	0.0%
Outcome: Intermediate	12	91.7%
Outcome: Patient Perspective	18	100%
Process	60	98.3%
Structure	3	66.7%
Total	119	95.0%

ⁱⁱ The three “Median Time” measures are evaluated on improvement, not performance. A performance standard could not be applied consistently across time measures.

ⁱⁱⁱ No performance for efficiency standards exists. These four Hospital OQR Program measures are excluded.

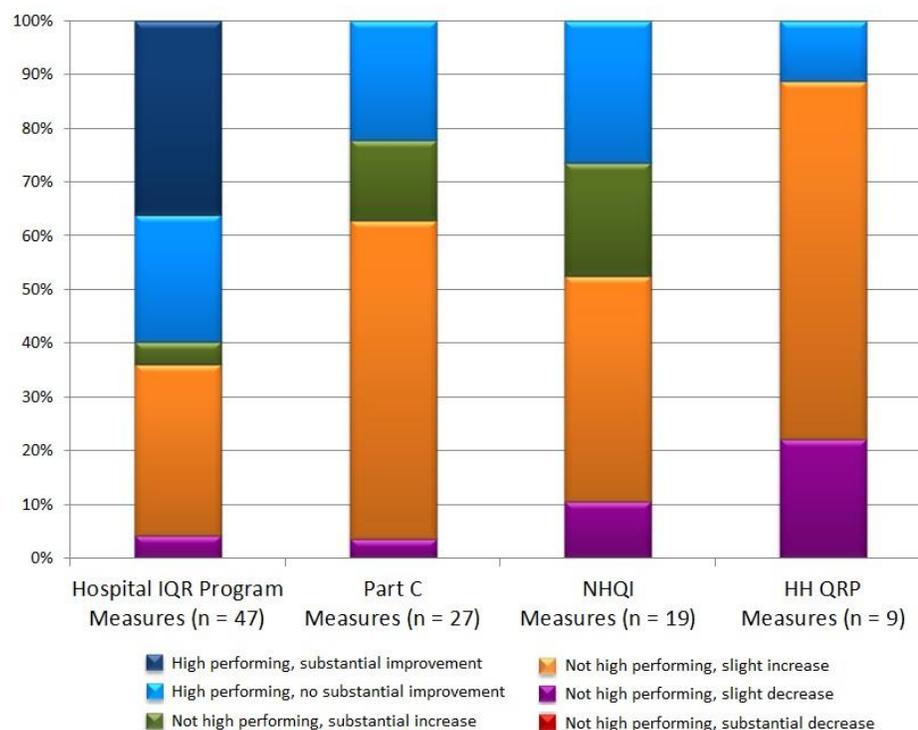
Performance and improvement for each quality measure are summarized by program in Table 7-5. High-performing measures (with rates of > 90 percent or < 5 percent over three consecutive measurement years) ranged from 0 percent reported in Part D and HH QRP to 60 percent of those reported for the Hospital IQR Program. Hospital programs (Hospital IQR Program and Hospital OQR Program) accounted for 76 percent (31 out of 41) of the high-performing measures and all of the high-performing and substantially increasing measures. Among ambulatory (Part C and Part D) programs, 11 percent (4 out of 35) of measures were high performing, another 23 percent (8 out of 35), substantially improved and 63 percent (22 out of 35) improved slightly. However, one measure, *Improving Bladder Control*,^{liii} had a slight decrease in performance. In the post-acute setting (HH QRP, NHQI, and End Stage Renal Disease Quality Incentive Program [ESRD QIP]), 20 percent (6 out of 30) were high performing, and another 20 percent (6 out of 30) demonstrated substantial improvement. No measures in the programs included in Table 7-5 substantially decreased at the national level.

Table 7-5: Counts and Percentages of the Quality Measures by Program by Levels of Performance and Improvement

High Performing?	Improvement Level	Hospital		Ambulatory		Post-Acute			Total
		Hospital IQR Program	Hospital OQR Program	Part C	Part D	HH QRP	NHQI	ESRD QIP	
Yes	Substantial	17 36.2%	2 28.6%	0	0	0	0	0	19 16.0%
	Not Substantial	11 23.4%	1 14.3%	4 14.8%	0	0	5 26.3%	1 50.0%	22 18.5%
Subtotal of High-Performing Measures		28 59.6%	3 42.9%	4 14.8%	0	0	5 26.3%	1 50.0%	41 34.5%
No	Substantial increase	2 4.3%	0	6 22.2%	2 25.0%	1 11.1%	4 21.1%	1 50.0%	16 13.4%
	Slight increase	15 31.9%	4 57.1%	16 59.3%	6 75.0%	6 66.7%	8 42.1%	0	56 47.1%
	Slight decrease	2 4.3%	0	1 3.7%	0	2 22.2%	2 10.5%	0	6 5.0%
	Substantial decrease	0	0	0	0	0	0	0	0
Subtotal of Non-High-Performing Measures		19 40.4%	4 57.1%	23 85.2%	8 100%	9 100%	14 73.7%	1 50.0%	78 65.5%
Total Number of Measures		47 100%	7 100%	27 100%	8 100%	9 100%	19 100%	2 100%	119 100%

Figure 7-2 presents the distribution of performance and improvement categories for programs with nine or more quality measures. The Hospital IQR Program had the largest percentage of measures that were high performing or substantially improving, and HH QRP had the fewest.

^{liii} NQF-endorsed title: *Urinary Incontinence Management in Older Adults—*a. Discussing urinary incontinence, b. Receiving urinary incontinence treatment (NQF #0030).

Figure 7-2: Distribution of Performance and Improvement of Measure Rates by Program

The measures reported among the different care settings varied by type (outcome, process, and structure), which may explain part of the variation in performance between settings.

Table 7-6 summarizes measures by type. Process measures are generally regarded as easier to implement and achieve higher performance than outcome measures, because they can be addressed by changing care delivery processes or provider behaviors. The structural measures included in this study involve participation in specialty registries that require greater resource investment for providers to use and may take additional time to implement. Outcome measures may take longer to improve, as they may be affected by factors outside the clinical setting.

- ◆ Process measures represent 83 percent (34 out of 41) of the high-performing measures assessed.
- ◆ Outcome measures represent 15 percent (6 out of 41) and structural measures represent 2 percent (one out of 41) of the high-performing measures assessed.
- ◆ For those measures that did not achieve high performance status, substantial increases are demonstrated in 18 percent (11 out of 60) of the process measures compared with 9 percent (5 out of 56) of the outcome measures.

Table 7-6: Counts and Percentages of Quality Measures by Level of Performance and Improvement by Measure Type

High Performing?	Improvement Level	Outcome	Process	Structure	Total
Yes	Substantial	0	19 31.7%	0	19 16.0%
	Not Substantial	6 10.7%	15 25.0%	1 33.3%	22 18.5%
Subtotal of High-Performing Measures		6 10.7%	34 56.7%	1 33.3%	41 34.5%
No	Substantial increase	5 8.9%	11 18.3%	0	16 13.4%
	Slight increase	40 71.4%	14 23.3%	1 33.3%	55 46.2%
	Slight decrease	5 8.9%	1 1.7%	1 33.3%	7 5.9%
	Substantial decrease	0	0	0	0
Subtotal of Non-High-Performing Measures		50 89.3%	26 43.3%	2 66.6%	78 65.5%
Total Number of Measures		56 100%	60 100%	3 100%	119 100%

Table 7-7 summarizes the distribution of measure types within each program.

- ◆ The post-acute setting has the largest percentage of outcome measures (73 percent, 22 out of 30).
- ◆ The hospital setting has the lowest percentage of outcome measures (30 percent, 16 out of 54).
- ◆ The hospital setting has the largest percentage of process measures (65 percent, 35 out of 54).
- ◆ The post-acute setting has the lowest percentage of process measures (27 percent, eight out of 30).
- ◆ The percentage of outcome and process measures for the ambulatory setting was similar (51 and 49 percent, 18 and 17 out of 35, respectively).

Table 7-7: Measure Types by Program

Program	Outcome	Process	Structure	Total
Hospital: Hospital IQR Program	16 28.6%	28 46.7%	3 100%	47 39.5%
Hospital: Hospital OQR Program	0	7 11.7%	0	7 5.9%
Subtotal of Hospital Setting Measures	16 28.6%	35 58.3%	3 100.0%	54 45.4%
Ambulatory: Part C	12 21.4%	15 25.0%	0	27 22.7%
Ambulatory: Part D	6 10.7%	2 3.3%	0	8 6.7%
Subtotal of Ambulatory Setting Measures	18 32.1%	17 28.3%	0	35 29.4%
Post-Acute: HH QRP	9 16.1%	0	0	9 7.6%
Post-Acute: NHQI	13 23.2%	6 10.0%	0	19 16.0%
Post-Acute: ESRD QIP	0	2 3.3%	0	2 1.7%
Subtotal of Post-Acute Setting Measures	22 39.3%	8 13.3%	0	30 25.3%
Total Number of Measures	56 100%	60 100%	3 100%	119 100%

As noted in Table 7-6, process measures were more likely to be high performing and improving. However, the reported performance of process and outcomes measures is not consistent between programs, as seen in Table 7-8.

Table 7-8: Measure Trends by Program and Measure Type

Measure Type and Program	High Performing		Not High Performing				Sub-total
	Substantial Improvement	No Substantial Improvement	Substantial Increase	Slight Increase	Slight Decrease	Substantial Decrease	
Outcome Measures	0	5	4	23	5	0	38
Hospital IQR Program	0	0	0	5	1	0	6
Hospital OQR Program	0	0	0	0	0	0	0
NHQI	0	4	1	6	2	0	13
HH QRP	0	0	1	6	2	0	9
Part C	0	2	1	4	0	0	7
Part D	0	0	1	2	0	0	3
ESRD	0	0	0	0	0	0	0
Patient Perspective	0	0	1	17	0	0	18
Hospital IQR Program	0	0	1	9	0	0	10
Hospital OQR Program	0	0	0	0	0	0	0
NHQI	0	0	0	0	0	0	0
HH QRP	0	0	0	0	0	0	0
Part C	0	0	0	5	0	0	5
Part D	0	0	0	3	0	0	3
ESRD	0	0	0	0	0	0	0

Table 7-8: Measure Trends by Program and Measure Type

Measure Type and Program	High Performing		Not High Performing				Sub-total
	Substantial Improvement	No Substantial Improvement	Substantial Increase	Slight Increase	Slight Decrease	Substantial Decrease	
Process Measures	19	15	11	14	1	0	60
Hospital IQR Program	17	10	1	0	0	0	28
Hospital OQR Program	2	1	0	4	0	0	7
NHQI	0	1	3	2	0	0	6
HH QRP	0	0	0	0	0	0	0
Part C	0	2	5	7	1	0	15
Part D	0	0	1	1	0	0	2
ESRD	0	1	1	0	0	0	2

Measure Trends by National Quality Strategy Priorities

A CMS focus in recent years has been on aligning quality measures with the National Quality Strategy (NQS) priorities. These priorities were published by the U.S. Department of Health and Human Services (HHS) in 2011⁷ and provide a framework for coordinating quality measure development, implementation, and maintenance efforts. The six measure domains aligned with the NQS priorities were identified as key areas on which to focus resources to achieve the aims of Better Care, Healthy People/Healthy Communities, and Affordable Care.

Table 7-9 illustrates performance on the 119 measures in this study according to the measure domains. Measures in the Safety (72 percent) and Effective Treatment (39 percent) domains have a disproportionate percentage of measures that are high performing. No measure saw a substantial decrease at the national level.

Table 7-9: Measure Trends by Measure Domains Between 2006 and 2012

High Performing?	Improvement Level	Safety	Patient Engagement	Care Coordination	Effective Treatment	Healthy Communities	Affordable Care
Yes	Substantial	9 50.0%	3 10.7%	0	7 12.5%	0	0
	Not Substantial	4 22.2%	2 7.1%	0	15 26.8%	0	1 100.0%
Subtotal of High-Performing Measures		13 72.2%	5 17.9%	0	22 39.3%	0	1 100.0%
No	Substantial increase	2 11.1%	2 7.1%	0	6 10.7%	6 54.5%	0
	Slight increase	3 16.7%	21 75.0%	5 100.0%	21 37.5%	5 45.5%	0
	Slight decrease	0	0	0	7 12.5%	0	0
	Substantial decrease	0	0	0	0	0	0
Subtotal of Non-High-Performing Measures		5 27.8%	23 82.1%	5 100%	34 60.7%	11 100%	0
Total Number of Measures		18 100%	28 100%	5 100%	56 100%	11 100%	1 100%

Objectives 1, 2, and 3—Trend Analyses by Care Settings

Performance Trends Among Hospital Setting Programs

Objectives 1 and 2—Performance and Improvement

- ◆ Hospital-based programs, the Hospital IQR Program and the Hospital OQR Program combined, account for 54 (45 percent) of the measures that met the inclusion criteria of three or more annual data points and nine reporting providers.
- ◆ Sixty-one percent (19 out of 31) of high-performing hospital measures continued to show substantial improvement throughout the study period.
- ◆ Each of the 28 measures for condition-specific processes met the criteria to be high performing, with the exception of *AMI-7a: Fibrinolytic Therapy Received Within 30 Minutes of Hospital Arrival* (NQF #0164), which was substantially increasing.
- ◆ Patient engagement measures demonstrated slight increases over the study period, as did the outcomes measures.

Success Stories

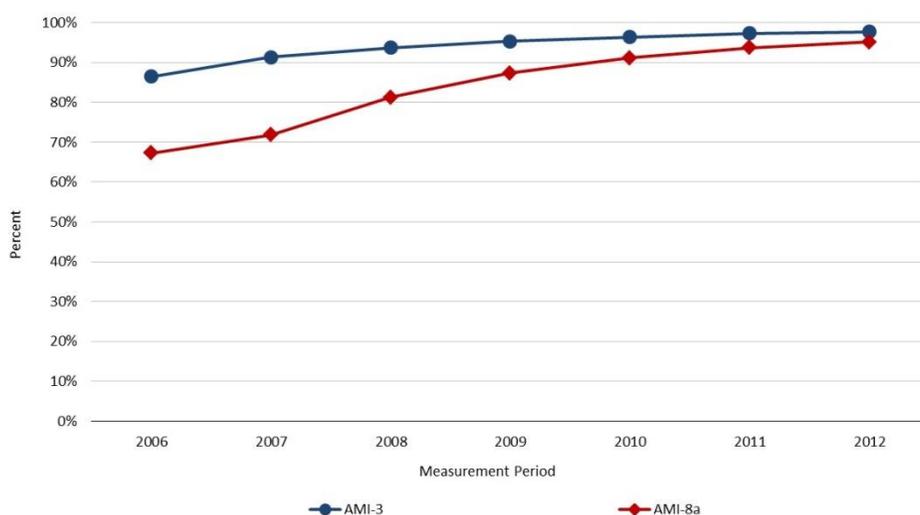
The Hospital IQR Program includes several condition-specific measure sets that have demonstrated substantial improvement in the delivery of inpatient care. The eight measures addressing appropriate care of acute myocardial infarction were either high performing or substantially increasing from 2006 to 2012. Two of the eight improving measures, AMI-3 and AMI-8a, are shown in Figure 7-3. Also, slight improvements are seen in the associated outcome measures *30-day Risk-Standardized Readmission*^{liv} and *30-day Risk-Standardized Mortality*^{lv} (19 to 18 percent and 16 to 15 percent, respectively).

Figure 7-3 and Table 7-10 present examples of AMI measures that have substantially improved. While both *AMI-3: AMI ACEI/ARB for LVSD*^{lvi} and *AMI-8a: Primary PCI Received Within 90 Minutes of Hospital Arrival* (NQF #0163) appear to be approaching their maximum, both exhibit noticeable increases between 2006 and 2012. CMS has subsequently removed these two measures from the Hospital IQR Program for Fiscal Year 2017 payment determination due to “topped-out” status.⁸

^{liv} NQF-endorsed title: *Hospital 30-day all-cause risk-standardized readmission rate (RSRR) following acute myocardial infarction (AMI) hospitalization* (NQF #0505).

^{lv} NQF-endorsed title: *Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute myocardial infarction (AMI) hospitalization for patients 18 and older* (NQF #0230)

^{lvi} NQF-endorsed title: *ACEI or ARB for left ventricular systolic dysfunction- Acute Myocardial Infarction (AMI) Patients* (NQF #0137).

Figure 7-3: Hospital IQR Program Acute Myocardial Infarction Measures With Substantial Improvement Between 2006 and 2012**Table 7-10: Measure Rates for Substantially Improving Hospital IQR Program AMI Process Measures**

Measure Name	2006	2007	2008	2009	2010	2011	2012
AMI-3: AMI ACEI/ARB for LVSD	86.4%	91.3%	93.7%	95.3%	96.3%	97.3%	97.6%
AMI-8a: Primary PCI Received Within 90 Minutes of Hospital Arrival	67.3%	71.8%	81.3%	87.4%	91.1%	93.7	95.1%

Three measures addressing appropriate treatment of heart failure patients were high performing and are presented in Figure 7-4 and Table 7-11. *HF-4: Adult Smoking Cessation Advice/Counseling* was removed from the Hospital IQR Program in 2011 (effective for Fiscal Year 2014 and subsequent payment determinations) when measure compliance reached a rate of 99 percent.⁹ *HF-1: Discharge Instructions*^{lvii} increased 25 percent between 2006 and 2012, while *HF-3: ACE Inhibitor or ARB for Left Ventricular Systolic Dysfunction (LVSD)* (NQF #0162) increased by 11 percent. *HF-1: Discharge Instructions* and *HF-3: ACE Inhibitor or ARB for Left Ventricular Systolic Dysfunction (LVSD)* were removed from the Hospital IQR Program in 2014 (effective for Fiscal Year 2016 and subsequent payment determinations).¹⁰ *HF-1: Discharge Instructions* was removed for several reasons, including loss of NQF endorsement status, weak correlation to outcomes, and factors related to patient understanding of the instructions.

^{lvii} Previously NQF-endorsed title: Heart Failure (HF): Detailed discharge instructions (NQF #0136).

Figure 7-4: Hospital IQR Program Heart Failure Measures With Substantial Improvement Between 2006 and 2012

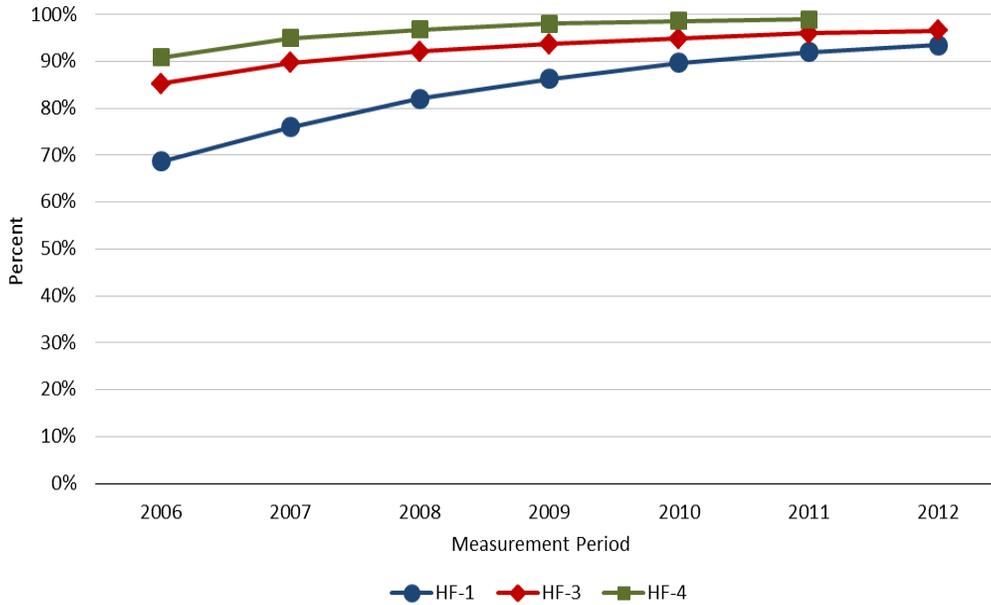


Table 7-11: Measure Rates for Substantially Improving Hospital IQR Program HF Process Measures

Measure Name	2006	2007	2008	2009	2010	2011	2012
HF-1: Discharge Instructions	68.7%	76.0%	82.0%	86.2%	89.7%	92.0%	93.5%
HF-3: ACE Inhibitor or ARB for Left Ventricular Systolic Dysfunction (LVSD)	85.2%	89.7%	92.1%	93.7%	94.9%	96.0%	96.6%
HF-4: Adult Smoking Cessation Advice/Counseling	90.8%	94.9%	96.8%	98.0%	98.6%	99.0%	R ^{lviii}

The four measures that are high performing for pneumonia are presented in Figure 7-5 and Table 7-12. Each of these four measures has been removed from the Hospital IQR Program: *PN-7: Influenza Vaccination Status*^{lix} and *PN-2: Pneumococcal Vaccination Status*^{lx} were removed in 2010 for Fiscal Year 2014 and subsequent payment determinations;¹¹ *PN-4: Adult Smoking Cessation Advice/Counseling* was removed in 2011 for Fiscal Year 2012 and subsequent payment determinations;⁹ and the chart-abstracted version of *PN-6: Appropriate Initial Antibiotic Selection*^{lxi} was removed in 2014 for Fiscal Year 2017 and subsequent payment

^{lviii} R=Removed First Quarter 2012

^{lix} Previously NQF-endorsed title: *Influenza vaccination* (NQF #0149).

^{lx} Previously NQF-endorsed title: *Pneumococcal vaccination* (NQF #0150).

^{lxi} NQF-endorsed title: *Initial antibiotic selection for community-acquired pneumonia (CAP) in immunocompetent patients* (NQF #0147).

determinations;^{lxii} however, the electronic version of this measure was retained. Before being removed, *PN-7: Influenza Vaccination Status* increased 26 percentage points, while *PN-2: Pneumococcal Vaccination Status* and *PN-4: Adult Smoking Cessation Advice/Counseling* increased 21 and 11 points, respectively. *PN-6: Appropriate Initial Antibiotic Selection* increased 9 percentage points.

Figure 7-5: Hospital IQR Program Pneumonia Measures With Substantial Improvement Between 2006 and 2012

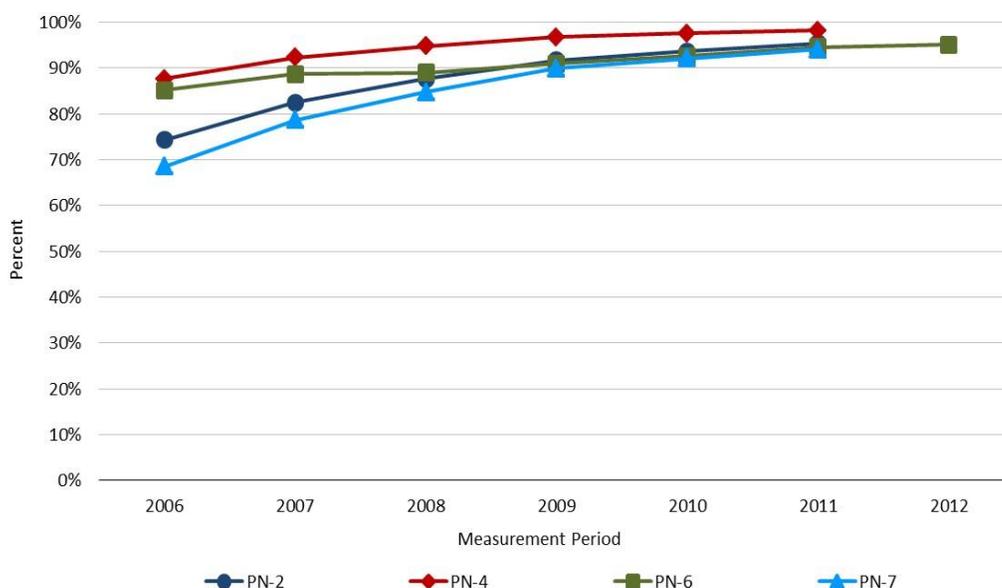


Table 7-12: Measure Rates for Substantially Improving Hospital IQR Program Pneumonia Process Measures

Measure Name	2006	2007	2008	2009	2010	2011	2012
PN-2: Pneumococcal Vaccination Status	74.3%	82.5%	87.6%	91.7%	93.7%	95.4%	R ^{lxii}
PN-4: Adult Smoking Cessation Advice/Counseling	87.6%	92.3%	94.8%	96.7%	97.6%	98.2%	R ^{lxiii}
PN-6: Appropriate Initial Antibiotic Selection	85.2%	88.7%	89.0%	90.9%	92.6%	94.6%	95.1% ^{lxiv}
PN-7: Influenza Vaccination Status	68.5%	78.7%	84.7%	89.9%	92.0%	94.1%	R ^{lxii}

Five Surgical Care Improvement Project (SCIP) measures improved substantially and were high performing during 2009 to 2012, as presented in Figure 7-6 and Table 7-13. The improvement of measures ranged from 21 percentage points for *SCIP-INF-3: Prophylactic*

^{lxii} R=Retired/Removed Fiscal Year 2014

^{lxiii} R=Retired/Removed Fiscal Year 2012

^{lxiv} R=To be Retired/Removed Fiscal Year 2017; electronic version of measure to be retained.

Antibiotics Discontinued Within 24 Hours After Surgery End Time (48 Hours for Cardiac Surgery) (NQF #0529) to 9 percent for *SCIP-INF-9: Surgery Patients Whose Urinary Catheters Were Removed on the First or Second Day After Surgery*.^{lxv} The measure rate for each measure is above 96 percent by the end of the study period (2012).

Figure 7-6: Hospital IQR Program Surgical Care Improvement Project Measures (Infection) With Substantial Improvement Between 2006 and 2012

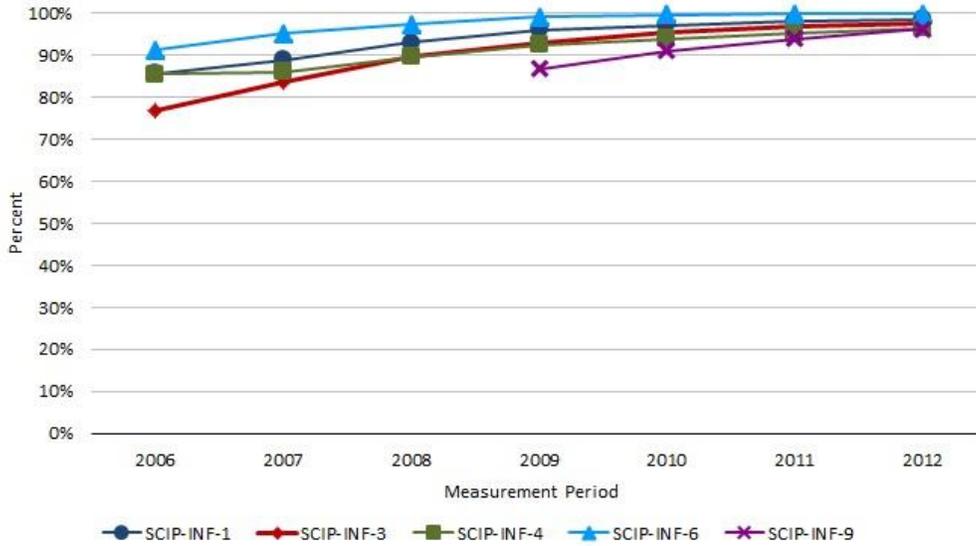


Table 7-13: Measure Rates for Substantially Improving Hospital IQR Program SCIP-INF Process Measures^{lxvi}

Measure Name	2006	2007	2008	2009	2010	2011	2012
SCIP-INF-1: Prophylactic Antibiotic Received Within One Hour Prior to Surgical Incision	85.8%	88.8%	93.0%	96.0%	97.2%	98.0%	98.5%
SCIP-INF-3: Prophylactic Antibiotics Discontinued Within 24 Hours After Surgery End Time (48 Hours for Cardiac Surgery)	77.0%	83.7%	89.6%	93.1%	95.5%	96.8%	97.6%
SCIP-INF-4: Cardiac Surgery Patients With Controlled Postoperative Blood Glucose	85.6%	86.1%	89.7%	92.5%	94.0%	95.2%	96.3%
SCIP-INF-6: Surgery Patients With Appropriate Hair Removal	91.3%	95.2%	97.2%	99.1%	99.6%	99.8%	99.8%
SCIP-INF-9: Surgery Patients Whose Urinary Catheters Were Removed on the First or Second Day After Surgery	N/A ^{lxvii}	N/A ^{lxvii}	N/A ^{lxvii}	86.8%	91.1%	93.9%	96.2%

^{lxv} NQF-endorsed title: *Urinary catheter removed on Postoperative Day 1 (POD1) or Postoperative Day 2 (POD2) with day of surgery being day zero* (NQF #0453).

^{lxvi} See Appendix i-4 for a hyperlink to a crosswalk of all measures in Table 7-13, including CMS Title, NQF Title, NQF I.D., and NQF endorsement status.

^{lxvii} Not Available—Indicates that the measure was not implemented in that year.

Opportunities for Improvement

In the Hospital IQR Program, “patient perspective” outcome measures from the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS, NQF #0166) Survey, registry participation, and readmission/mortality rates demonstrated slight but consistent improvement over the study period.

- ◆ Within the HCAHPS survey, rates for measures related to Nurse and Doctor Communication, Staff Responsiveness, Pain Management, Pharmacy Communication, Cleanliness, Quietness, Discharge Instructions, Would Recommend, and Overall Ratings improved between 0.5–1.0 percent per year, increasing by 3 to 5 percent from 2008 to 2012.
- ◆ Cardiac Surgery Registry participation increased from 89 percent in 2009 to 95 percent in 2012. Stroke Care Registry and Nursing Care Registry participation remained at 45 percent from 2010 to 2012.
- ◆ Thirty-day readmission rates for AMI, heart failure, and pneumonia discharges improved slightly from 2009 to 2012, ranging from 17 to 23 percent.

Within the Hospital OQR Program, measures *OP-1: Median Time to Fibrinolysis*,^{lxviii} *OP-2: Fibrinolytic Therapy Received Within 30 Minutes of ED Arrival* (NQF #0288), and *OP-3: Median Time to Transfer to Another Facility for Acute Coronary Intervention* (NQF #0290) demonstrated slight improvements over the study period.

Objective 3—Performance and Improvement of Measure Rates by Provider Characteristics

Hospital IQR Program Process Measures

Performance and improvement of quality measures for the Hospital IQR Program providers are presented by selected hospital characteristics in Table 7-14. Rural hospitals, those with less than 50 beds, and Safety Net hospitals had fewer high-performing and fewer substantially increasing measures, and government-owned hospitals reported fewer high-performing measures. Teaching hospitals had more high-performing measures than non-teaching hospitals (93 percent and 79 percent, respectively), but more non-teaching hospitals than teaching hospitals were substantially or slightly increasing over the study period (18 percent versus 7 percent).

^{lxviii} Previously NQF-endorsed title: *Median Time to Fibrinolysis* (NQF #0287)

Table 7-14: Percentages of the 28 Hospital IQR Program Process Measures by Levels of Performance and Improvement by Provider Characteristics

Provider Characteristics		High Performing		Not High Performing			
		Substantial Improvement	No Substantial Improvement	Substantial Increase	Slight Increase	Slight Decrease	Substantial Decrease
National		60.7%	35.7%	3.6%	0.0%	0.0%	0.0%
Urbanicity	Urban	39.3%	50.0%	3.6%	3.6%	3.6%	0.0%
	Rural	7.1% ^{lxix}	25.0%	7.1%	53.6%	7.1%	0.0%
Ownership	For-Profit	28.6%	60.7%	7.1%	3.6%	0.0%	0.0%
	Not-for-Profit	28.6%	57.1%	7.1%	3.6%	3.6%	0.0%
	Government	7.1%	42.9%	10.7%	35.7%	3.6%	0.0%
Teaching Status	Non-Teaching	21.4%	57.1%	7.1%	10.7%	3.6%	0.0%
	Teaching	57.1%	35.7%	3.6%	3.6%	0.0%	0.0%
Safety Net Hospital	Safety Net	11.1%	48.1%	14.8%	22.2%	3.7%	0.0%
	Non-Safety Net	25.9%	59.3%	7.4%	3.7%	3.7%	0.0%
Bed Size	<50	7.1%	17.9%	0.0%	64.3%	10.7%	0.0%
	50–200	25.0%	57.1%	10.7%	3.6%	3.6%	0.0%
	>200	67.9%	28.6%	0.0%	3.6%	0.0%	0.0%

Hospital OQR Program Process Measures

Provider characteristics are summarized for the seven Hospital OQR Program measures in Table 7-15. Because the number of Hospital OQR Program measures is small, counts are presented in the tables rather than percentages. Although not enough Hospital OQR Program measures were available to evaluate trends, there appears, as with the Hospital IQR Program, to be a greater proportion of high-performing, substantially increasing measures among urban hospitals, teaching hospitals, and those with more than 200 beds.

Table 7-15: Performance and Improvement of the Quality Measure Rates by Provider Characteristics for Seven Hospital OQR Program Process Measures

Provider Characteristics		High Performing		Not High Performing			
		Substantial Improvement	No Substantial Improvement	Substantial Increase	Slight Increase	Slight Decrease	Substantial Decrease
National		2	1	0	4	0	0
Urbanicity	Urban	1	2	0	4	0	0
	Rural	0	3	0	4	0	0

^{lxix} It is possible to find different percentages of measures showing different levels of improvement at the national level when compared to that of provider level for two reasons. First, the percentages for the provider characteristics are not a simple subdivision of the percentage at the national level. Second, the analytic approach is accounting for the variability across providers.

Table 7-15: Performance and Improvement of the Quality Measure Rates by Provider Characteristics for Seven Hospital OQR Program Process Measures

Provider Characteristics		High Performing		Not High Performing			
		Substantial Improvement	No Substantial Improvement	Substantial Increase	Slight Increase	Slight Decrease	Substantial Decrease
Ownership	For-Profit	0	3	0	4	0	0
	Not-for-Profit	0	3	0	4	0	0
	Government	0	3	0	2	2	0
Teaching Status	Non-Teaching	0	3	0	4	0	0
	Teaching	1	2	0	2	2	0
Safety Net Hospital	Safety Net	0	3	0	4	0	0
	Non-Safety Net	0	3	0	4	0	0
Bed Size	<50	0	3	0	3	1	0
	50–200	0	3	0	3	1	0
	>200	2	1	0	4	0	0

Performance Trends Among Ambulatory Setting Programs

CMS initiated 10 programs applied to the ambulatory setting. Two programs had measures with the required three annual data points available for analysis: Medicare Part C and Medicare Part D Programs. Although PQRS and Medicare Electronic Prescribing Incentive Program (eRx Incentive Program) measures also had the required three annual data points, provider-level data did not meet the inclusion criterion of being publically available. Aggregate trends for these programs were reported in the *CMS 2012 Reporting Experience Including Trends (2007–2013)*.¹³ Effective 2014, measures from PQRS and Accountable Care Organizations (ACOs) are reported on Physician Compare. These publically reported measures will meet inclusion criteria for future reports.

Objectives 1 and 2—Performance and Improvement

- ◆ Thirty-five quality measures met the inclusion criteria for this study, including 17 process measures and 18 outcome measures for Medicare Parts C and D collectively.
- ◆ Four measures (11 percent) were high performing, and eight (23 percent) were substantially increasing, with 16 (46 percent) having a slight increase and one with a slight decrease between 2007 and 2013 (Appendix 7-6 and Appendix 7-7).

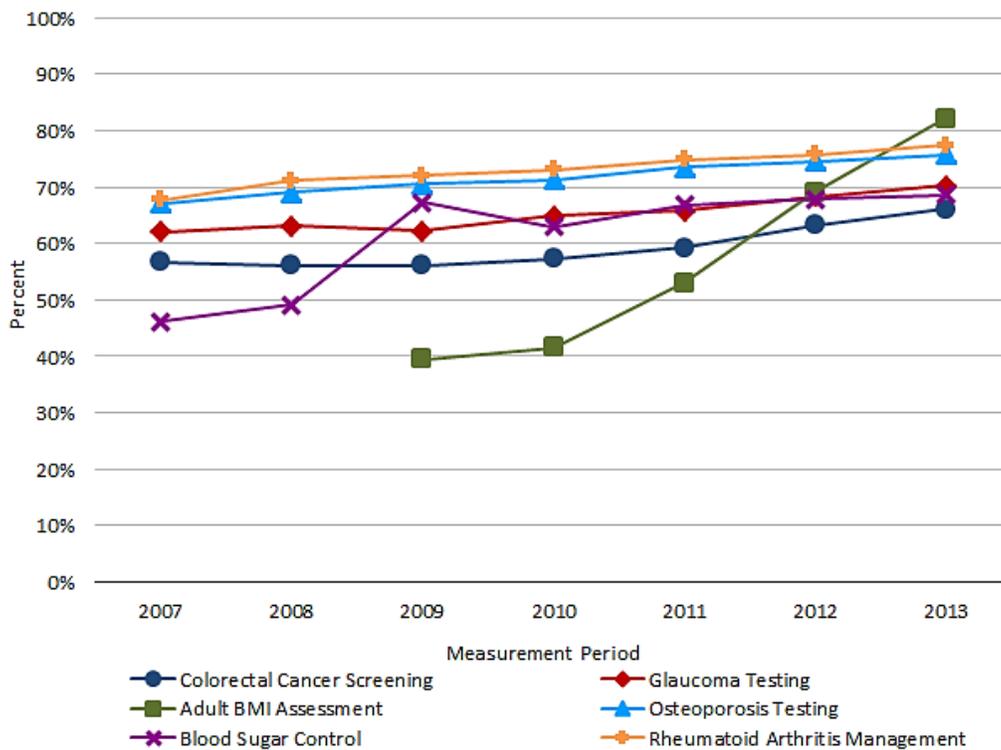
Success Stories

Four of the 27 (15 percent) Part C measures were high performing, and six (22 percent) demonstrated substantial improvement. Of the eight Part D measures, none were high performing, and two demonstrated substantial improvement over the study period. The quality measures studied included a large number of outcomes measures (44 percent of Part C and 75

percent of Part D). Among the Part C measures, process measures *Cholesterol Screening for Patients With Diabetes* (NQF #1780), *Cholesterol Screening for Patients With Heart Disease*,^{lxx} and *Kidney Function Screening for Members With Diabetes*^{lxxi} were high performing, along with the outcome measure *Adults' Access to Preventive/Ambulatory Health Services (65+)*.

Figure 7-7 and Table 7-16 present six Part C screening measures that showed substantial improvement over the study period. The improvement in the measures ranged from 43 percentage points for *Adult BMI Assessment* and 22 percentage points for *Plan Members With Diabetes Whose Blood Sugar Is Under Control*^{lxxii} to 8 percentage points for *Glaucoma Testing*.

Figure 7-7: Medicare Part C Measures With Substantial Change Between 2007 and 2013



^{lxx} NQF-endorsed title: *Ischemic Vascular Disease (IVD): Complete Lipid Profile and LDL-C Control <100 mg/dL* (NQF #0075).

^{lxxi} NQF-endorsed title: *Comprehensive Diabetes Care: Medical Attention for Nephropathy* (NQF #0062).

^{lxxii} NQF-endorsed title: *Comprehensive Diabetes Care: Hemoglobin A1c (HbA1c) Poor Control (>9.0%)* (NQF #0059).

Table 7-16: Rates for Medicare Part C Measures With Substantial Change^{lxxiii}

Measure Name	2007	2008	2009	2010	2011	2012	2013
Colorectal Cancer Screening	56.8%	56.1%	56.1%	57.3%	59.3%	63.2%	66.1%
Glaucoma Testing	62.1%	63.0%	62.1%	64.9%	65.7%	68.4%	70.3%
Adult BMI Assessment	N/A ^{lxxiv}	N/A ^{lxxiv}	39.5%	41.5%	53.0%	69.0%	82.2%
Osteoporosis Testing in Older Women	67.1%	69.1%	70.6%	71.3%	73.5%	74.6%	75.7%
Plan Members With Diabetes Whose Blood Sugar Is Under Control	46.2%	49.1%	67.3%	62.9%	66.8%	67.8%	68.5%
Rheumatoid Arthritis Management	67.7%	71.1%	72.0%	73.1%	74.9%	75.7%	77.3%

Of the eight Part D quality measures, two showed substantial improvement. *Use of High-Risk Medications in the Elderly* (NQF #0022) decreased from 23 percent in 2008 to 6 percent in 2012, and *Taking Blood Pressure Medication*^{lxxv} improved from 73 percent in 2010 to 77 percent in 2012 (see Appendix 7-9 and 7-10 for details on each measure).

Opportunities for Improvement

- ◆ Sixty-three percent of the quality measures had slight improvements. These included seven Part C Medicare Advantage and Prescription Drug Plan (MA & PDP) CAHPS measures of patient perspective of care and outcome measures, such as *Improving or Maintaining Mental Health*, *Improving or Maintaining Physical Health*, *Controlling Blood Pressure* (NQF #0018), and *Plan Members With Diabetes Whose Cholesterol Is Under Control*.^{lxxvi}
- ◆ The Part D measures with slight improvement included five CAHPS patient perspective measures and *Appropriate Treatment of Hypertension for Diabetics*.^{lxxvii} These results suggest that although the process measures for screening patients with diabetes for complications is improving, the corresponding outcome measures for treatment of those complications is lagging.
- ◆ Other screening and health measures with slight improvements included *Reducing the Risk of Falling*,^{lxxviii} *Monitoring Physical Activity*,^{lxxix} *Breast Cancer Screening for Women 52-69*,^{lxxx} *Eye Exam to Check for Damage from Diabetes*,^{lxxxi} *Osteoporosis Management in Women Who Had a Fracture* (NQF #0053), *Pneumonia Vaccine*,^{lxxxii} and *Annual Flu Vaccine*.^{lxxxiii}

^{lxxiii} See Appendix i-4 for a hyperlink to a crosswalk of all measures in Table 7-16, including CMS Title, NQF Title, NQF I.D., and NQF endorsement status.

^{lxxiv} Not Available—Indicates that the measure was not implemented in that year.

^{lxxv} NQF-endorsed title: *Proportion of Days Covered (PDC): 5 Rates by Therapeutic Category* (NQF #0541).

^{lxxvi} NQF-endorsed title: *Comprehensive Diabetes Care: LDL-C Control <100 mg/dL* (NQF #0064).

^{lxxvii} NQF-endorsed title: *Diabetes: Appropriate Treatment of Hypertension* (NQF #0546).

^{lxxviii} NQF-endorsed title: *Fall Risk Management (FRM)* (NQF #0035).

^{lxxix} NQF-endorsed title: *Physical Activity in Older Adults (PAO)* (NQF #0029).

^{lxxx} Previously NQF-endorsed title: *Breast Cancer Screening* (NQF #0031).

^{lxxxi} NQF-endorsed title: *Comprehensive Diabetes Care: Eye Exam (retinal) performed* (NQF #0055).

^{lxxxii} NQF-endorsed title: *Pneumococcal Vaccination Status for Older Adults (PNU)* (NQF #0043).

^{lxxxiii} Previously NQF-endorsed title: *Flu Shot for Older Adults* (NQF #0040).

Objective 3—Performance and Improvement of Measure Rates by Provider Characteristics for Part C Measures

Table 7-17 presents results for Part C measures by plan characteristics for ownership, plan type,^{lxxxiv} and special needs plan (SNP)^{lxxxv} type. Urbanicity is tabulated at the level of individual Part C members. Unlike hospital-based providers, high performing Part C measures did not differ by urbanicity. Not-for-profit plans reported more high-performing measures than for-profit plans (15 percent and 4 percent, respectively). Two-thirds of measures for dual eligible SNPs met the performance and substantial improvement standards as compared to 24 percent of measures for Medicare Advantage plans.

Table 7-17: Performance and Improvement on the Quality Measure Rates by Plan Characteristics for 27 Part C Measures

Plan Characteristic	High Performing		Not High Performing				
	Substantial Improvement	No Substantial Improvement	Substantial Increase	Slight Increase	Slight Decrease	Substantial Decrease	
National	0.0%	14.8%	22.2%	59.3%	3.7%	0.0%	
Urbanicity ^{lxxxvi}	Urban	0.0%	7.4%	22.2%	66.7%	3.7%	0.0%
	Rural	0.0%	3.7%	37.0%	51.9%	7.4%	0.0%
Ownership	For-Profit	0.0%	3.7%	25.9%	70.4%	0.0%	0.0%
	Not-for-Profit	0.0%	14.8%	25.9%	51.9%	7.4%	0.0%
Plan Type	HMO/HMO POS	0.0%	14.8%	22.2%	59.3%	3.7%	0.0%
	Local PPO	0.0%	3.7%	22.2%	63.0%	11.1%	0.0%
	Regional PPO	0.0%	3.7%	25.9%	51.9%	18.5%	0.0%
	Private Fee-for-Service	0.0%	3.7%	29.6%	51.9%	14.8%	0.0%
	Other	0.0%	11.1%	22.2%	48.1%	18.5%	0.0%
SNP Type	Medicare Advantage	4.8%	4.8%	14.3%	71.4%	4.8%	0.0%
	Chronic Condition	0.0%	9.5%	42.9%	19.0%	28.6%	0.0%
	Dual Eligible	0.0%	14.3%	52.4%	28.6%	4.8%	0.0%
	Institutional	0.0%	14.3%	14.3%	14.3%	57.1%	0.0%

^{lxxxiv} Plan types include health maintenance organization and health maintenance organization point of service combinations (HMO/HMO POS), local and regional preferred provider organizations (Local PPO, Regional PPO), private fee-for-service, and plans that do not fit into the preceding categories (Other).

^{lxxxv} Medicare special needs plans (SNPs) limit membership to people with specific diseases or characteristics and tailor their benefits, provider choices, and drug formularies to best meet the specific needs of the groups they serve. Qualifying individuals can enroll in four types of plans: a Medicare Advantage, plan, a plan that specializes in individuals with chronic conditions (Chronic Conditions), a plan that specializes in individuals who are eligible for Medicare and Medicaid (Dual Eligible), or a plan that specializes in individuals who are institutionalized (Institutional).

^{lxxxvi} Urbanicity is tabulated across individual Part C members.

Performance Trends Among Post-Acute Setting Programs

Objectives 1 and 2—Performance and Improvement

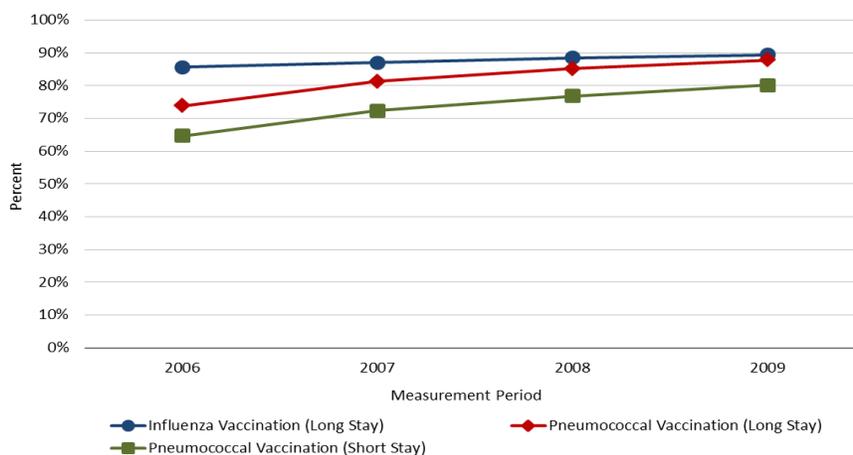
- ◆ Post-acute programs that were assessed include NHQI, HH QRP, and ESRD QIP. Other post-acute programs (e.g., hospice) were not included because their data did not meet the data inclusion criteria for this study (see Methods section).
- ◆ NHQI data from the MDS 2.0 dataset was collected from 2006 to 2009. In 2010, the MDS was updated to address changes in nursing home care, resident characteristics, and advances in resident assessment methods. Because versions 2.0 and 3.0 have substantial differences, this 2015 Impact Report examines MDS 2.0. Results from the updated MDS 3.0 will be presented in the 2018 Impact Report.
- ◆ The OASIS-B1 quality measures for the HH QRP were updated in 2010; 2006 to 2009 trends are reported here. Results from the updated OASIS-C will be presented in the 2018 Impact Report.

Success Stories

High-performing NHQI measures included low rates of *Residents Who Have Moderate to Severe Pain (Long Stay)*,^{lxxxvii} *Residents Who Spend Most of Their Time in a Bed or in a Chair (Long Stay)*,^{lxxxviii} *Residents Who Were Physically Restrained (Long Stay)* (NQF #0687), *Residents with Delirium (Short Stay)*,^{lxxxix} and *Low-risk Residents With Pressure Ulcers (Long Stay)*. None of the substantially improving NHQI measures achieved high performance standards during the study period; three of them are shown in Figure 7-8 and Table 7-18.

Within the HH QRP program, *Improvement in Ambulation/Locomotion* (NQF #0167) reports increased from 41 percent in 2006 to 46 percent in 2009. Both of the ESRD measures were classified as either high performing or substantially improving.

Figure 7-8: NHQI Process Quality Measures Showing Substantial Improvement Between 2006 and 2009



^{lxxxvii} NQF-endorsed title: *Percent of Residents Who Self-Report Moderate to Severe Pain (Long-Stay)* (NQF #0677).

^{lxxxviii} Previously NQF-endorsed title: *Residents who spent most of their time in bed or in a chair in their room during the 7-day assessment period* (NQF 0194).

^{lxxxix} Previously NQF-endorsed title: *Recently hospitalized residents with symptoms of delirium (risk-adjusted)* (NQF 0185).

Table 7-18: Quality Measure Rates for Substantially Improving NHQI Process Measures^{xc}

Measure Name	2006	2007	2008	2009
Influenza Vaccination (Long Stay)	85.7%	87.0%	88.5%	89.5%
Pneumococcal Vaccination (Long Stay)	73.8%	81.3%	85.2%	87.9%
Pneumococcal Vaccination (Short Stay)	64.6%	72.3%	76.9%	80.2%

Opportunities for Improvement

- ◆ Several outcome measures had slightly increased rates, but rates decreased for *Improvement in Dyspnea^{xci}* and *Patients Able to Live in the Community at Discharge*.
- ◆ Several clinical outcomes measures also demonstrated slightly increased rates in NHQI, related to catheter use and UTIs, moderate/severe pain management, weight loss, and functional status.

Objective 3—Performance and Improvement of Measure Rates by Provider Characteristics

Quality measure results by selected nursing home provider characteristics are shown in Table 7-19. There is less variation in performance by provider characteristic than for hospital providers (Table 7-14). For example, 32 percent of measures for rural nursing homes were high performing, compared with 21 percent of urban nursing homes. Performance rates were similar between facilities with different ownership types and in chain versus independent facilities, although chain facilities tended to report more substantially increasing measure rates than independent facilities (42 percent versus 21 percent). The homes with the fewest beds (up to 64 beds) reported more high-performing measures (32 percent versus 21 percent in Quartiles 2–4) and fewer substantially increasing measures (11 percent versus 26–47 percent) than nursing homes with large bed sizes.

The NHQI results included the one measure whose rates substantially decreased at the national level in this study: *Long-Stay Low-Risk Residents Lose Control of Bowels/Bladder* (NQF #0685). This rate substantially decreased for facilities that had not-for-profit ownership, that were in Quartiles 3 and 4 of bed size, and were in Quartile 3 of nursing hours.

^{xc} See Appendix i-4 for a hyperlink to a crosswalk of all measures in Table 7-18, including CMS Title, NQF Title, NQF I.D., and NQF endorsement status.

^{xci} Previously NQF-endorsed title: *Improvement in Dyspnea* (NQF #0179).

Table 7-19: Percentages of the 19 NHQI Measures by Levels of Performance and Improvement by Provider Characteristics^{xcii}

Provider Characteristics		High Performing		Not High Performing			
		Substantial Improvement	No Substantial Improvement	Substantial Increase	Slight Increase	Slight Decrease	Substantial Decrease
National		0.0%	26.3%	21.1%	42.1%	10.5%	0.0%
Urbanicity	Urban	5.3%	15.8%	42.1%	21.1%	15.8%	0.0%
	Rural	10.5%	21.1%	26.3%	26.3%	15.8%	0.0%
Ownership	For-Profit	5.3%	15.8%	36.8%	31.6%	10.5%	0.0%
	Not-for-Profit	5.3%	21.1%	26.3%	31.6%	10.5%	5.3%
	Government	10.5%	21.1%	15.8%	36.8%	15.8%	0.0%
Affiliation	Chain	5.3%	15.8%	42.1%	31.6%	5.3%	0.0%
	Independent	5.3%	15.8%	21.1%	42.1%	15.8%	0.0%
Bed Size	Quartile 1	10.5%	21.1%	10.5%	42.1%	15.8%	0.0%
	Quartile 2	5.3%	15.8%	26.3%	36.8%	15.8%	0.0%
	Quartile 3	5.3%	15.8%	47.4%	21.1%	5.3%	5.3%
	Quartile 4	5.3%	15.8%	47.4%	21.1%	5.3%	5.3%
Nursing Hours	Quartile 1	5.3%	15.8%	26.3%	36.8%	15.8%	0.0%
	Quartile 2	5.3%	15.8%	42.1%	21.1%	15.8%	0.0%
	Quartile 3	5.3%	15.8%	47.4%	26.3%	0.0%	5.3%
	Quartile 4	5.3%	10.5%	42.1%	31.6%	10.5%	0.0%

Objective 4—Patient Characteristics and Disparities

Objective 4: Quantitatively assess disparities in initial measure rates by age, sex, race and ethnicity and determine whether disparities diminish over the study period.

In this section, the hospital, ambulatory, and post-acute settings are examined first, followed by a presentation of overall results. Within each subsection, the descriptive analysis of performance and improvement is presented first, followed by results of the disparities analysis by race and ethnicity, followed by illustrative examples of changes in disparities on individual measures.

The descriptive analysis was performed for 83 measures from four CMS quality reporting programs: Hospital IQR Program (n=28), Medicare Part C (n=27), NHQI (n=19), and HH QRP (n=9). The disparities analysis was conducted as described in the Methods section. Age, sex, and race or race/ethnicity information were available for 104 of the 119 CMS quality measures analyzed elsewhere in this chapter. Ethnicity was collected separately from race for 59 of the measures. For each of the settings (hospitals, ambulatory, and post-acute), disparities by race and ethnicity are shown; disparities by age and sex in each setting are presented and briefly discussed in the concluding section. Results for each measure are included in Appendices 7-12 through 7-21.

^{xcii} The quartiles for Bed Size (from Quartile 1 to Quartile 4, respectively) are: ≤ 63 beds, 64-99 beds, 100-129 beds, and ≥ 130 beds. The quartiles for RN Hours per resident per day (from Quartile 1 to Quartile 4, respectively) are: ≤ 0.61 hours, 0.62 – 0.79 hours, 0.80 – 0.98 hours, ≥ 0.99 hours.

Hospital Settings

At the national level (all patients), 27 of the 28 hospital measures in the Hospital IQR Program were classified as high performing (Table 7-20). This pattern largely held when the measures were examined by age, sex, and ethnicity. There were notable differences by race, however. While 82 percent of the measures were classified as high performing among American Indian/Alaska Natives and 89 percent among Black patients, over 96 percent were high performing among White and Asian patients. Whether this represents differences in access to the well-performing large urban hospitals rather than differential treatment within facilities is a question worth examining.

Table 7-20: Percentages of the 28 Hospital IQR Program Process Measures by Levels of Performance and Improvement by Age, Sex, and Race/Ethnicity

Subpopulation		High Performing		Not High Performing			
		Substantial Improvement	No Substantial Improvement	Substantial Increase	Slight Increase	Slight Decrease	Substantial Decrease
National		60.7%	35.7%	3.6%	0.0%	0.0%	0.0%
Age	18–64	59.3%	37.0%	3.7%	0.0%	0.0%	0.0%
	65–84	57.1%	39.3%	3.6%	0.0%	0.0%	0.0%
	85+	57.1%	39.3%	3.6%	0.0%	0.0%	0.0%
Sex	Female	60.7%	35.7%	3.6%	0.0%	0.0%	0.0%
	Male	60.7%	35.7%	3.6%	0.0%	0.0%	0.0%
Race ^{xciii}	AIAN	50.0%	32.1%	17.9%	0.0%	0.0%	0.0%
	Asian	67.9%	28.6%	3.6%	0.0%	0.0%	0.0%
	Black	60.7%	28.6%	10.7%	0.0%	0.0%	0.0%
	NHPI	57.1%	32.1%	7.1%	0.0%	3.6%	0.0%
	White	57.1%	39.3%	3.6%	0.0%	0.0%	0.0%
	Other	53.6%	35.7%	10.7%	0.0%	0.0%	0.0%
Ethnicity	Non-Hispanic	57.1%	39.3%	3.6%	0.0%	0.0%	0.0%
	Hispanic	57.1%	32.1%	10.7%	0.0%	0.0%	0.0%

A disparities analysis was carried out on 27 measures from the Hospital IQR Program (one measure used in the descriptive analysis was dropped due to data reliability problems). Providers reported race and ethnicity separately for the Hospital IQR Program; the reference categories for the disparity determination were White for race and non-Hispanic for ethnicity. There were race disparities in 12 of the measures, and 11 of the 12 diminish between 2006 and 2012. There were disparities by ethnicity for 10 of the measures, and each of these disparities diminished over time.

Figure 7-9, Table 7-21, Figure 7-10, and Table 7-22 illustrate two success stories in narrowing disparities. For the process measures *Smoking Cessation Advice/Counseling (PN-4)* and *Primary*

^{xciii} The race categories were American Indian/Alaska Native (AIAN), Asian, Black, Native Hawaiian/Pacific Islander (NHPI), White, and Other/Unknown.

PCI Received Within 90 Minutes of Hospital Arrival (AMI-8A), the percentage of patients from all groups who received the standard of care increased rapidly. The quality of care for most of the other groups improved more rapidly than for Whites, narrowing disparities. Particularly striking are the disappearance of a White-Asian disparity in *Smoking Cessation Advice/Counseling (PN-4)* over a four-year period and the near-disappearance of a White-Black disparity in *Primary PCI Received Within 90 Minutes of Hospital Arrival (AMI-8A)* over a slightly longer period. While the measures were not high performing for all groups by the final three years of the study period, each group appeared to be on track to achieve high-performing status.

Figure 7-9: Smoking Cessation Advice/Counseling (PN-4) by Race Between 2006 and 2012

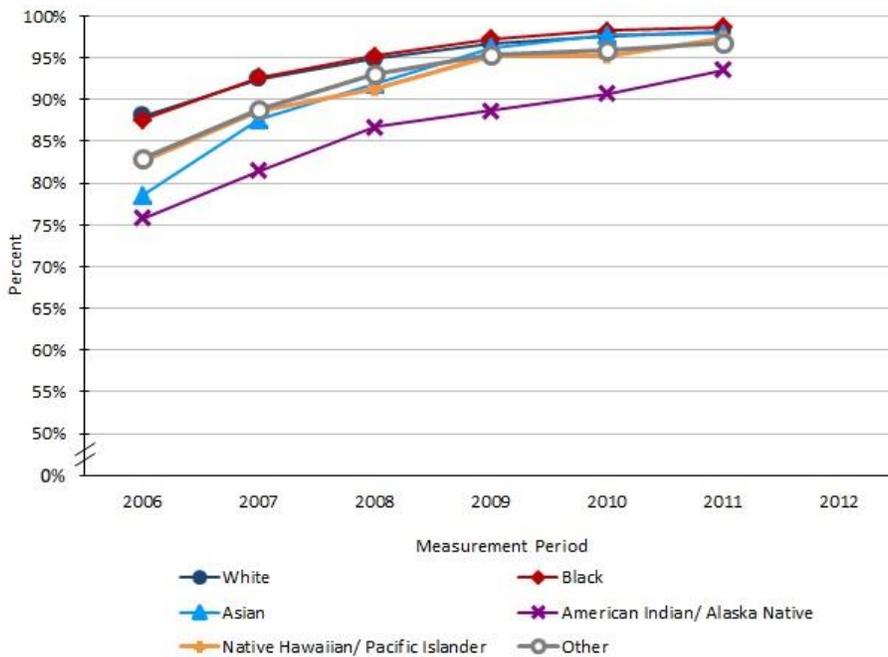


Table 7-21: Smoking Cessation Advice/Counseling (PN-4) by Race Between 2006 and 2012

Subgroup Name	2006	2007	2008	2009	2010	2011	2012
White	88.1%	92.5%	94.9%	96.8%	97.6%	98.2%	R ^{xciv}
Black	87.6%	92.7%	95.3%	97.3%	98.2%	98.7%	R ^{xciv}
Asian	78.6%	87.6%	91.9%	96.2%	97.7%	97.9%	R ^{xciv}
American Indian/Alaska Native	75.8%	81.5%	86.7%	88.7%	90.7%	93.6%	R ^{xciv}
Native Hawaiian/Pacific Islander	82.5%	88.7%	91.4%	95.2%	95.3%	97.4%	R ^{xciv}
Other	83.0%	88.9%	93.0%	95.5%	95.9%	96.8%	R ^{xciv}

^{xciv} R=Retired/Removed First Quarter 2012

Figure 7-10: Primary PCI Received Within 90 Minutes of Hospital Arrival (AMI-8A) by Race Between 2006 and 2012

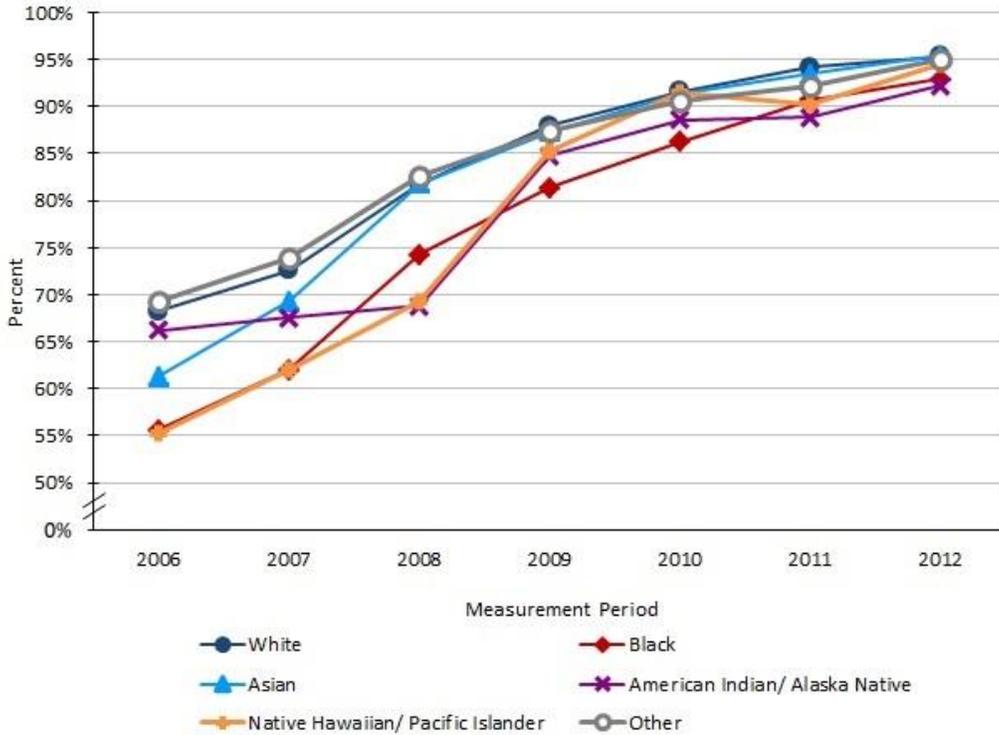


Table 7-22: Primary PCI Received Within 90 Minutes of Hospital Arrival (AMI-8A) by Race Between 2006 and 2012

Subgroup Name	2006	2007	2008	2009	2010	2011	2012
White	68.3%	72.6%	81.9%	87.9%	91.6%	94.2%	95.3%
Black	55.6%	62.0%	74.3%	81.4%	86.2%	90.7%	92.9%
Asian	61.3%	69.3%	81.8%	87.3%	91.4%	93.5%	95.4%
American Indian/Alaska Native	66.2%	67.6%	68.8%	84.7%	88.5%	88.9%	92.2%
Native Hawaiian/Pacific Islander	55.3%	62.0%	69.3%	85.3%	91.4%	90.3%	94.4%
Other	69.3%	73.9%	82.7%	87.4%	90.5%	92.3%	95.0%

Ambulatory Settings

The data on ambulatory settings come from the Medicare Part C Program. Part C had few high-performing measures (Table 7-23). However, 52 percent of the measures were either high performing or substantially improving for the Hispanic group. Moreover, greater than half of these measures were improving slightly or substantially for all groups, ranging from a low of 69 percent for “Other” race/ethnicity to a high of 96 percent for ages 18–64.

Table 7-23: Percentages of the 27 Medicare Part C Measures by Levels of Performance and Improvement by Age, Sex, and Race/Ethnicity

Subpopulation	High Performing		Not High Performing				
	Substantial Improvement	No Substantial Improvement	Substantial Increase	Slight Increase	Slight Decrease	Substantial Decrease	
National	0.0%	14.8%	22.2%	59.3%	3.7%	0.0%	
Age	18–64	0.0%	0.0%	36.0%	60.0%	4.0%	0.0%
	65–84	0.0%	14.8%	33.3%	48.1%	3.7%	0.0%
	85+	0.0%	8.3%	16.7%	66.7%	8.3%	0.0%
Sex	Female	0.0%	14.8%	18.5%	66.7%	0.0%	0.0%
	Male	0.0%	12.0%	20.0%	64.0%	4.0%	0.0%
Race/ Ethnicity ^{xcv}	AIAN	0.0%	0.0%	7.7%	76.9%	15.4%	0.0%
	Asian/PI	0.0%	14.8%	22.2%	59.3%	3.7%	0.0%
	Black	0.0%	7.4%	22.2%	66.7%	3.7%	0.0%
	Other	0.0%	0.0%	0.0%	69.2%	30.8%	0.0%
	Non-Hispanic White	0.0%	3.7%	7.4%	81.5%	7.4%	0.0%
	Hispanic	3.7%	7.4%	40.7%	48.1%	0.0%	0.0%

For Part C, providers reported race/ethnicity as a single variable. The reference group for the analysis was non-Hispanic Whites. Hispanics were treated as one of several race/ethnicity groups, so there was no separate analysis of disparities by ethnicity. Disparities were found for one or more groups in 19 out of 27 Part C measures. Greater than half (14) of these disparities were found to be diminishing over time.

Post-Acute Settings

Data from the NHQI and the HH QRP are included in the post-acute setting analyses. Measures from other post-acute care programs will be incorporated into future reports as data meeting criteria for inclusion become available.

Nursing Homes (NHQI)

Patient-level results for age, sex, race/ethnicity, and ethnicity are presented for the NHQI measures in Table 7-24. There were few differences between age, sex, race, and ethnicity groups. However, the 18 and younger age group had more high-performing and/or substantially improving measures compared to the 19–64 age group (63 versus 37 percent).

^{xcv} The race/ethnicity categories are: American Indian/Alaska Native (AIAN), Asian/Pacific Islander (Asian/PI), Black, Multicultural/Other/Unknown, Hispanic, and non-Hispanic White.

Table 7-24: Percentage of the NHQI Quality Measures by Performance Levels and by Age, Sex, and Race/Ethnicity and Ethnicity Alone

Subpopulation		High Performing		Not High Performing			
		Substantial Improvement	No Substantial Improvement	Substantial Increase	Slight Increase	Slight Decrease	Substantial Decrease
National		0.0%	26.3%	21.1%	42.1%	10.5%	0.0%
Age	≤18	5.3%	36.8%	21.1%	31.6%	0.0%	5.3%
	19–64	0.0%	15.8%	21.1%	42.1%	21.1%	0.0%
	65–84	0.0%	26.3%	15.8%	42.1%	15.8%	0.0%
	85+	5.3%	31.6%	10.5%	36.8%	15.8%	0.0%
Sex	Female	0.0%	31.6%	15.8%	36.8%	15.8%	0.0%
	Male	0.0%	26.3%	15.8%	42.1%	15.8%	0.0%
Race/ Ethnicity ^{xcvi}	AIAN	0.0%	26.3%	15.8%	47.4%	10.5%	0.0%
	Asian/PI	5.3%	26.3%	21.1%	31.6%	15.8%	0.0%
	Black	0.0%	15.8%	21.1%	47.4%	15.8%	0.0%
	Hispanic	5.3%	15.8%	15.8%	36.8%	26.3%	0.0%
	White	0.0%	26.3%	15.8%	42.1%	15.8%	0.0%
	Other	0.0%	21.1%	15.8%	57.9%	5.3%	0.0%
Ethnicity	Non-Hispanic	0.0%	26.3%	15.8%	42.1%	15.8%	0.0%
	Hispanic	5.3%	15.8%	15.8%	36.8%	26.3%	0.0%

NHQI reports race and ethnicity together in a race/ethnicity variable but the research team also created a separate ethnicity variable. Disparities between non-Hispanic Whites and non-Whites were found for four of 19 measures. The NHQI measures identified with racial disparities were *Influenza Vaccination (Short Stay)* (NQF #0680), *Pneumococcal Vaccination (Short Stay)* (NQF #0682), *Influenza Vaccination (Long Stay)* (NQF #0681), and *Pneumococcal Vaccination (Long Stay)* (NQF #0683). Each of these disparities diminished over time. Disparities between Hispanics and non-Hispanics were found for three of 19 measures. These disparities were undiminished.

The overall improvement in *Pneumococcal Vaccination (Short Stay)* and the narrowing of disparities over a four-year period was not dramatic; however, this was a success story for non-Hispanic Blacks, in particular, even though a Black-White disparity remains (Figure 7-11 and Table 7-25). In 2006, the rate at which Blacks received the standard of care was 17 percentage points less than Whites. By 2009, there had been a 23 percentage point increase in Blacks receiving the standard of care, and the disparity between Blacks and Whites had narrowed to 11 percent.

^{xcvi} The race/ethnicity categories were American Indian/Alaska Native (AIAN), Asian/Pacific Islander (Asian/PI), non-Hispanic Black, Hispanic, non-Hispanic White, and Other/Unknown. Ethnicity is created by grouping those identified as Hispanic in the Hispanic group and all others in the non-Hispanic group.

Figure 7-11: Pneumococcal Vaccination (Short Stay) by Race Between 2006 and 2009

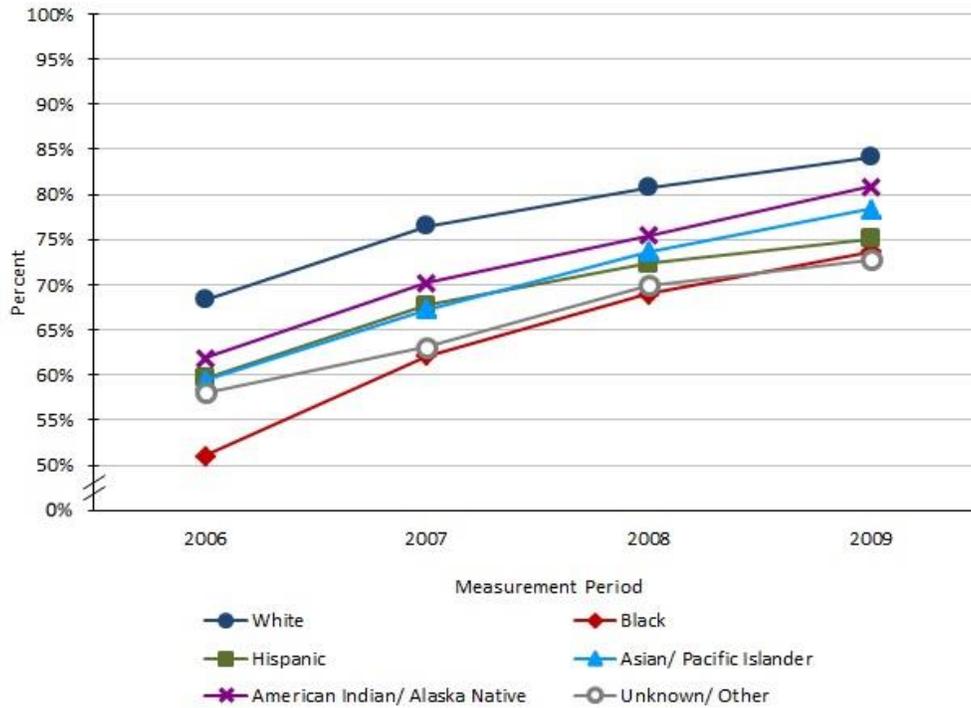


Table 7-25: Pneumococcal Vaccination (Short Stay) by Race Between 2006 and 2009

Subgroup Name	2006	2007	2008	2009
White	68.4%	76.5%	80.8%	84.2%
Black	51.0%	62.1%	68.9%	73.7%
Hispanic	59.6%	67.7%	72.3%	75.0%
Asian/Pacific Islander	59.4%	67.3%	73.7%	78.3%
American Indian/Alaska Native	61.9%	70.2%	75.5%	80.9%
Unknown/Other	58.0%	63.1%	70.0%	72.8%

Home Health Measures

The nine measures analyzed in this subsection were drawn from HH QRP. Few differences in performance and improvement on the quality measures were observed across demographic groups on these home health measures. No measures were classified as high performing, but most demographic groups followed the national pattern, with seven of nine measures improving slightly or substantially. Five measures were improving for Hispanics.

HH QRP reports both race/ethnicity and ethnicity alone. Disparities were found for four of nine measures by race. No disparities were found between Hispanics and non-Hispanics. None of the detected disparities were diminishing.

Cross-Program Trends

Differences in performance and improvement by age, sex, and race and ethnicity were analyzed for programs that collected the necessary data. The analyses at setting and program levels, however, showed few pronounced differences between national patterns and those of demographic groups, and most of those differences were for Blacks, Hispanics, and other racial/ethnic minorities. An exception was higher performance for the under 18 age group on nursing home measures, suggesting either that pediatric facilities deliver higher quality care or that the standards of care are easier to meet for younger patients.

With respect to disparities, the picture is more complex (Table 7-26). For the 104 quality measures, nine disparities were identified by sex, compared to 50 by race and ethnicity. The way disparities are reported actually understates the contrast between the two categories; a disparity between males and females on a single measure is counted as one disparity; disparities between Whites and Blacks, Asians/Pacific Islanders, and American Indians/Alaska Natives on a single measure would also be counted as one disparity.

While disparities by race and ethnicity were most numerous, they were also the most likely to have improved over the study period, with nearly one-third of the initial disparities diminishing or disappearing. The large number of disparities indicates that continued attention to disparities in the quality of care by race should remain a priority, but the observed reductions in disparities also shows that the healthcare system can make progress toward eliminating them.

Table 7-26: Counts and Percentages of Quality Measures by Levels of Disparities by Age, Sex, Race or Race/Ethnicity, and Ethnicity Alone

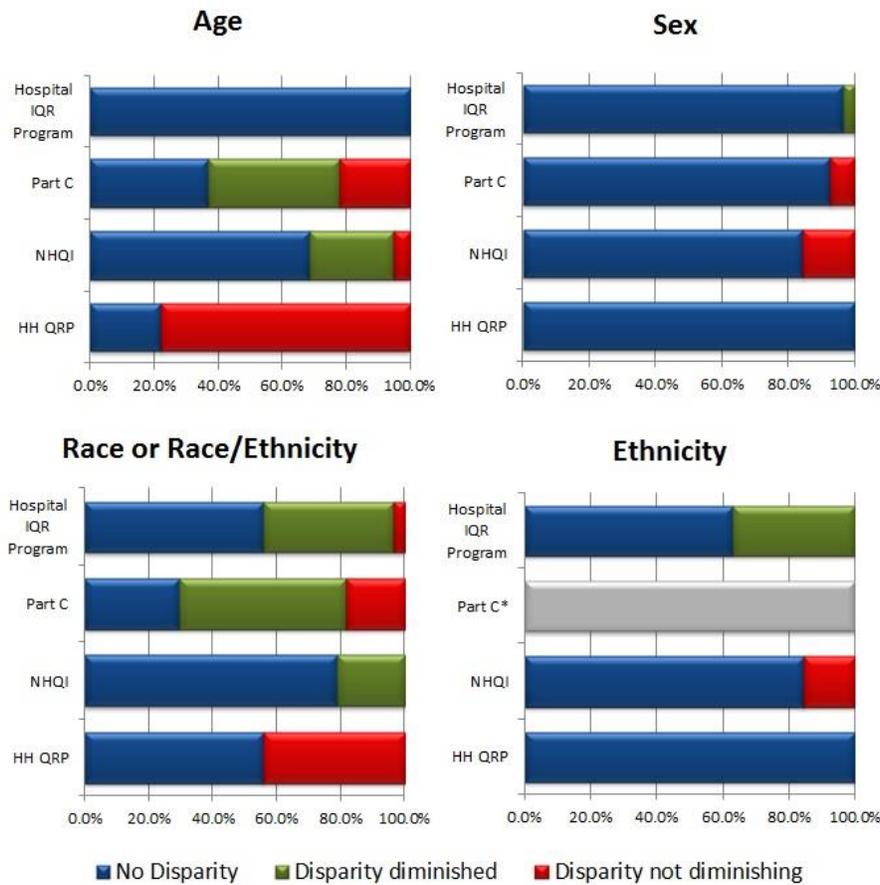
Group	No Disparity	Disparity Improving	Disparity Not Improving	Total
Age	66	16	22	104
	63.5%	15.4%	21.2%	100.0%
Sex	95	2	7	104
	91.3%	1.9%	6.7	100.0%
Race or Race/Ethnicity	54	33	17	104
	51.9%	31.7%	16.3%	100.0%
Ethnicity	46	10	3	59
	78.0%	16.9%	5.1%	100.0%

Figure 7-12 illustrates graphically both Table 7-26 and the preceding analyses by setting of care. Figure 7-12 draws attention to unevenness in the distribution of disparities and diminishing disparities across categories and programs. The evidence appears to suggest that certain programs have been more effective than others in reducing disparities. Marked differences across programs and settings in the measures studied, the characteristics of the data, the profile of the patients, etc., make comparisons across programs and settings problematic.

- ◆ No age disparities were detected among the Hospital IQR Program quality measures, but age disparities were present for seven of nine of HH QRP (home health) measures.
- ◆ Few sex disparities were found. Sex disparities in the Hospital IQR Program diminished, while none in Nursing Home Quality Improvement Program did.

- ◆ Disparities diminished for 11 of the 19 Medicare Part C (ambulatory) measures for which a disparity was identified, but none of the seven identified disparities on HH QRP measures diminished.
- ◆ Disparities by race and ethnicity were found in each program.
- ◆ As shown in Figure 7-12, NHQI measures showed the fewest disparities (four race/ethnicity disparities and three disparities by ethnicity alone), while Medicare Part C measures had race/ethnicity disparities for 19 of 27 measures, and the Hospital IQR Program had disparities in 12 measures by race and 10 measures by ethnicity.
- ◆ Seventy-four percent of the Part C race/ethnicity disparities diminished over the study period, however, and all but one of the Hospital IQR Program race/ethnicity disparities diminished over time.

Figure 7-12: Disparities Identified and Diminished for Age, Sex, Race, and Ethnicity



* Ethnicity data not collected separately for Medicare Part C

Discussion

CMS introduced quality measures in the 1990s with the goal of supporting nationwide quality improvement and monitoring the progress toward this goal. This study aims to quantify the magnitude of improvements achieved using CMS quality measures in a comprehensive national assessment, and provides a high-level analysis of performance on the quality measures across the different CMS programs with an interest in identifying trends over time, areas of initial disparity among different demographic groups, and assessing changes in disparity over time. The results presented here provide evidence of substantial success in the multi-year, national initiative to promote quality improvement and ensure that each patient receives the best care possible.

In this study, the research team used a novel approach to quantify the magnitude of shifts in the performance of a subset of CMS quality measures used in national programs. This study extends the 2012 Impact Analysis by examining two dimensions:

- ◆ “Performance” is defined by the level of the current measure rate relative to a high-performance threshold, where the threshold depends on whether the desirable target rate is high or low for the measure.
- ◆ “Improvement” classified the magnitude of the change in a rate for a measure over time using Cohen’s *d* metric of effect size.

Measures classified as non-high performing and substantially increasing are expected to eventually reach the target to become high performing.

Programs demonstrated some improvement in performance in 89 to 100 percent of their measures, and approximately 95 percent of the 119 measures analyzed for this report demonstrated improved performance to a degree. None of the measures analyzed demonstrated a substantial decrease in performance. These results suggest that on a national level, providers are engaged in the quality improvement process and improving patient care.

An important theme revealed by these analyses is that providers are performing well on measures of processes over which they have the most control. Seventy-five percent of process measures assessed in this study were classified as either high performing or substantially improving. Process measures that were not either high performing or substantially increasing over the study period were those that are in general not under providers’ complete control. For example, OP-3 (*Median Time to Transfer to Another Facility for Acute Coronary Intervention*) is dependent on the availability of another facility for admission and the availability and timeliness of independent transport providers. Process measures, such as administration of vaccination, providing screenings and management of chronic conditions, and monitoring activity and fall risks, require patient consent and changes in patient behavior. These measures were implemented to encourage providers to cooperate, find creative solutions, enlist community support, enhance patient engagement, and improve patient communication. The dependence of these measures on the actions of patients and others is likely to lead to the rates for the measures improving more slowly than rates for measures that can be improved by providers’ changes alone.

In contrast to the process measures, outcome measures (of which 20 percent were high performing or substantially improving) have several other determinants of success that are beyond the complete control of the provider. Performance of clinical outcome measures typically lags behind the performance of associated process measures. For instance, acute myocardial infarction process measures are applied to nearly all inpatients (Figure 7-3), but the *30-day Risk-Standardized Readmission* and *30-day Risk-Standardized Mortality* rates for AMI discharges demonstrated slight improvement. The readmission and mortality outcome rates may not be improving as fast as for the AMI process measures, in part, because they require collaboration and coordination with other providers, the patient, and sometimes the patient's family, which is more difficult than changing one's own internal processes. For example, hospital readmission measures require the hospitals to work with other providers across the continuum of care and to enhance their patient engagement strategies. Heart failure and pneumonia measures follow the same pattern. Rates of patient perspective measures, derived from patient surveys of their hospital experience, e.g., *Communication with Doctors*, *Responsiveness of Hospital Staff*, *Cleanliness*,^{xcvii} are also not improving as rapidly as the rates for process measures.

The CMS Quality Strategy focuses on aligning quality measures with National Quality Strategy (NQS) priorities. These priorities were published by the U.S. Department of Health and Human Services (HHS) in 2011 and provide a framework for coordinating quality measure development, implementation, and maintenance efforts. Although the quality measures included in this study were established prior to 2011, four of the six priorities were well represented, with 56 effective treatment, 18 safety, 28 patient engagement, 11 healthy communities, five care coordination measures, and one affordable care measures. Seventy-two percent of the safety measures were classified as high performing, indicating that providers have successfully implemented processes designed to improve patient safety, particularly the Surgical Care Improvement Project (SCIP) measures. Twenty-two (39 percent) of the 56 effective treatment measures were also high performing, with another six (11 percent) substantially improving. These include clinical guidelines for AMI, heart failure and pneumonia treatment, as well as screening for complications among patients with diabetes. Among the patient engagement measures, those that addressed provider behavior (*Discharge Information*, *Adult Smoking Cessation Counseling*) were high performing, but those that assessed patient opinions were not high performing and demonstrated slight increases over the study period.

Performance and improvement of measure rates varied by program as well. Hospital quality measures were high performing in 57 percent of the measures in this study, compared with 11 percent among the ambulatory and 20 percent among the post-acute settings. As noted above, many process measures in the ambulatory settings require collaboration and coordination with other providers, the patient, and sometimes the patient's family. For example, Medicare Part C process measures addressed vaccinations and preventive screening procedures, e.g., *Pneumonia Vaccine*, *Annual Flu Vaccine*, *Breast Cancer Screening*, *Eye Exam*, which require patient compliance.

^{xcvii} *Communication with Doctors*, *Responsiveness of Hospital Staff*, and *Cleanliness* are all components of HCAHPS (NQF #0166).

The research team found differences in performance rates of quality measures based on facility characteristics (Objective 3), including urbanicity, ownership, teaching and safety net status, and bed size in hospital settings, and based on urbanicity, ownership, affiliation, and bed size in nursing homes. These facilities may have differed in their access to resources, incentives to improve, and different quality improvement cultures. The facility characteristics associated with high-performing quality measures differed between hospital and nursing home settings. Among hospitals, a greater number of measures were high performing or substantially improving in hospitals that were urban, non-government, teaching, non-safety net, or large (> 200 beds) than in other hospitals. In contrast, among nursing homes, the urban, non-government, and large facilities had fewer high-performing measures than their counterparts. The reason for this finding is unclear and deserves further study.

The research team found a number of age, sex, race and ethnicity disparities in provider performance on the CMS quality measures, but also found that many of the disparities are diminishing (Objective 4). A modified version of the methodology developed by AHRQ for the 2013 *National Healthcare Disparities Report* was employed to identify disparities and to detect those that have diminished over time. Relatively few differences in performance and improvement on the quality measures were found across demographic groups overall or by setting of care.

For both disparities and provider performance and improvement on the quality measures, however, the most striking findings regard race and ethnicity rather than age and sex. Disparities in performance and improvement on quality measures were found in 22 percent of measures by ethnicity and 48 percent of measures by race and race/ethnicity. Disparities in performance and improvement on quality measures were found in 9 percent of measures by sex and 37 percent of measures by age. An interesting finding is that disparities diminished over time by 77 percent for ethnicity and 66 percent by race and race/ethnicity as compared to only 22 and 42 percent, respectively, for sex and age. Disparities diminished disproportionately, particularly for ethnicity; a larger percentage of existing disparities closed for ethnicity than for age, sex, or race and race/ethnicity.

There were also notable variations in disparities by programs and care settings. The Hospital IQR Program and Part C had the most measures with identified disparities, and most of those disparities improved over the reporting period. NHQI, HH QRP, and Part C had a number of age disparities, while the Hospital IQR Program had none. Whether the detected differences in disparities across programs and settings is a function of the particular measures being examined or the effort being made to eliminate disparities within the settings cannot be determined without further data and analysis, but the findings suggest which areas most need attention.

As noted above, the methods used by AHRQ in its 2013 *National Healthcare Disparities Report*¹ were adapted for use in this report. The similarity in methods between the AHRQ report and this research may give rise to questions about whether findings from the two are similar as well. Too many differences exist between the studies to permit a detailed comparison of results; however, in general, the differences between the two reports include differences in methods, differences in the sets of measures used, and differences in the data sources used. Despite these differences, the findings presented here are broadly consistent with those reported by AHRQ. The two studies generally agree that disparities in healthcare quality exist, particularly by race and

ethnicity; that although healthcare quality has been improving for all groups on many measures, disparities persist; and that while not all disparities are diminishing (e.g., American Indian/Alaska Native and Native Hawaiian/Pacific Islanders), many are. There is also agreement that while progress in reducing disparities has been substantial, much more work to decrease these disparity gaps is needed.

Conclusions

Key Findings

Ninety-five percent of the publicly reported measure rates across seven quality reporting programs showed improvement during the study period (2006–2012). Measures that address clinical guidelines for patient care (process measures) were most likely to be high performing (i.e., measure rates exceeding 90 percent in the three final years for which data were available). Process measures may have a limited lifespan since performance benchmarks are more rapidly achieved. Few measures that addressed clinical outcomes were high performing, though most demonstrated consistent improvements over the study period.

Provider and facility characteristics reflective of available resources (e.g., practice size, size of population served, and location) appear to be associated with increased provider participation in quality reporting programs and higher performance on quality measures. Nursing homes are an exception to this finding as urban, non-government, and large nursing home facilities had fewer high-performing measure rates than their counterparts did. Widespread race and ethnicity disparities existed at the beginning of the study period in 2006, and many had diminished by the end of the study period in 2012. However, it is important to note that generalizability of disparity findings is limited due to lack of standardized data collection.

Approximately 35 percent of all measures were classified as high performing, i.e., measure rates exceeding 90 percent or less than or equal to 5 percent in the three final years for which data were available. Further improvement on these measures may provide marginal returns in terms of impact on patient outcomes. CMS has started to develop objective criteria related to measure retirement, specific to individual reporting programs; however, an approach could be considered related to measure retirement that is sensitive to identified disparities.

Widespread race and ethnicity disparities that existed at the beginning of the study period in 2006 were much less pronounced in 2012; however, disparities persist across select programs, settings, and demographic groups. While lack of comparability in how race and ethnicity are assessed across settings limits generalizability, measure rates for Hispanics, Blacks, and Asians improved the most, and measure rates for American Indian/Alaska Native and Native Hawaiian/Pacific Islanders improved the least.

Actions to Consider

The results of this study suggest improvements in patient care across all CMS reporting programs and demographic groups occurred during the study period. However, since providers' performance on outcome measures improves more slowly, process measures may provide more

real-time information on the progress of quality improvement efforts to both CMS and providers. This will allow CMS and providers to make faster adjustments to program policies and quality intervention efforts.

Action to Consider: Reserve the development of process measures to those processes of care that link directly to patient outcomes and in which significant variation in performance exists across providers. Conversely, retire existing process measures that do not meet these criteria. Although providers’ performance improves more slowly for outcome measures than process measures, the latter can provide important real-time information on the progress of quality improvement efforts to both CMS and providers. This will allow CMS and providers to make contemporaneous adjustments to program policies and quality intervention efforts.

Although CMS has initiated the phasing out of process-of-care measures whose projected future impact on patient outcomes is limited, approximately 57 percent of process-of-care measures were classified as high performing (i.e., measure rates exceeding 90 percent in the three final years for which data were available). Further improvement of these measures may provide marginal returns in terms of impact on patient outcomes. CMS has started to develop objective criteria related to measure retirement, specific to individual reporting programs; however, an approach could be considered related to measure retirement that is sensitive to identified disparities.

Action to Consider: Develop clinically valid performance thresholds for high-performing measures that take into account any identified disparities, above which individual provider rates would be publicly reported as “exceeding performance standards.” Once measure rates exceed established performance thresholds, providers can focus resources on quality measures that do not meet performance thresholds.

Action to Consider: Develop standardized criteria in collaboration with NQF to retire quality measures. The criteria should include an assessment of national provider performance that accounts for overall performance means, distribution of measure rates, and disparity analysis. These criteria will objectively and uniformly inform the decision to retire measures from a reporting program. Measure developers could apply the criteria during the NQF comprehensive review, which is currently on a three-year cycle.

Provider characteristics appear to be associated with performance on measures across reporting programs. In particular, characteristics reflective of resource availability (e.g., provider size, location) may be indicative of higher performance. The QIN-QIOs have expertise and resources that can assist smaller providers in analyzing processes, systems, and procedures to improve overall quality and to reduce disparities.¹⁴

Action to Consider: Consider whether and to what extent the Quality Innovation Network-Quality Improvement Organizations (QIN-QIOs), as part of the 11th Scope of Work, can be directed to provide more focused technical assistance to smaller hospitals and physician practices and larger nursing homes.

Forty-two percent of detected disparities in the quality of care by age diminished over the course of the study period. Sixty-six percent of disparities by race or race/ethnicity also diminished, as did 77 percent of disparities by ethnicity alone. While progress in reducing disparities in these areas has been substantial, much more work remains. Another step to achieve the goal of eliminating racial and ethnic disparities is to increase the availability, quality, and use of data to improve the health of minority populations. In this study, variability in the collection of race and ethnicity data categories in support of the disparity analyses was particularly challenging.

Action to Consider: Standardize race and ethnicity data collection across CMS quality reporting programs to facilitate disparity analysis. Use existing standards developed by HHS, as required by section 4302 of the Affordable Care Act.^{5, 6} Consider including additional demographic characteristics such as disability status, educational level, and English proficiency. Implementation of standardized data elements will increase generalizability of analyses across reporting programs.

Action to Consider: Promote transparency and decrease the disparities in measure performance by publicly reporting and systematically monitoring the measure rates by race and ethnicity for all publicly reported measures. These measure results can be compared within peer groups of providers serving disproportionately large minority and underserved populations.

Reference List

- (1) Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Lawrence Erlbaum Associates; 1988.
- (2) Ross JS, Bernheim SM, Lin Z et al. Based on key measures, care quality for Medicare enrollees at safety-net and non-safety-net hospitals was almost equal. *Health Aff (Millwood)*. 2012;31(8):1739-1748.
- (3) Centers for Medicare & Medicaid Services. *CMS Quality Strategy - 2013 - Beyond*. Baltimore, MD: Centers for Medicare & Medicaid Services; November 18, 2013.
- (4) U.S. Department of Health and Human Services. *2013 National Healthcare Disparities Report*. Rockville, MD: Agency for Healthcare Research and Quality. May 2014.
- (5) Office of the Assistant Secretary for Planning and Evaluation. Implementation Guidance on Data Collection Standards for Race, Ethnicity, Sex, Primary Language, and Disability Status. <http://aspe.hhs.gov/datacncl/standards/ACA/4302/index.shtml>. Published October 2011. Accessed November 20, 2014.
- (6) Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity. *Fed Regist*. 2014;62(210):58782-58790.
- (7) U.S. Department of Health and Human Services. *Report to Congress: National Strategy for Quality Improvement in Health Care*. Washington, DC: U.S. Department of Health and Human Services; March 20, 2011.
- (8) Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long Term Care Hospital Prospective Payment System and Fiscal Year 2015 Rates Final Rule. *Fed Regist*. 2014;79(163):50208.
- (9) Medicare Program; Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long-Term Care Hospital Prospective Payment System and FY 2012 Rates; Hospitals' FTE Resident Caps for Graduate Medical Education Payment; Final Rule. *Fed Regist*. 2011;76(160):51611
- (10) Medicare Program; Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long Term Care; Hospital Prospective Payment System and Fiscal Year 2014 Rates; Quality Reporting Requirements for Specific Providers; Hospital Conditions of Participation; Payment Policies Related to Patient Status; Final Rule. *Fed Regist*. 2013;78(160):50782.
- (11) Medicare Program; Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long-Term Care Hospital Prospective Payment System Changes and FY2011 Rates; Provider Agreements and Supplier Approvals; and Hospital Conditions of Participation for Rehabilitation and Respiratory Care Services; Medicaid Program: Accreditation for Providers of Inpatient Psychiatric Services. *Fed Regist*. 2010;75(157):50211.

- (12) Medicare Program; Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long-Term Care Hospital Prospective Payment System and Fiscal Year 2015 Rates; Quality Reporting Requirements for Specific Providers; Reasonable Compensation Equivalents for Physician Services in Excluded Hospitals and Certain Teaching Hospitals; Provider Administrative Appeals and Judicial Review; Enforcement Provisions for Organ Transplant Centers; and Electronic Health Record (EHR) Incentive Program; Final Rule. *Fed Regist.* 2014;79(163):50208.
- (13) Centers for Medicare & Medicaid Services. *2012 Reporting Experience Including Trends (2007-2013): Physician Quality Reporting System and Electronic Prescribing (eRx) Incentive Program*. Baltimore, MD: Centers for Medicare & Medicaid Services; Mar 14, 2014.
- (14) Shea M, Triantis M, Passmore S, Oh E, Thomas S. *Hidden in Plain Sight: The Role of CMS Quality Improvement Organizations and the Elimination of Racial and Ethnic Health Disparities*. Columbia, MD: Delmarva Foundation for Medical Care. June 16, 2014.

Chapter 8—Measure Relationships: Hospital Process Measures and Patient Outcomes



Questions on Effectiveness

1. Are changes in performance rates for clinical process-of-care measures associated with changes in patient outcomes?
2. Are changes in performance rates for clinical process-of-care measures associated with changes in health outcomes of the target population?

Abstract

Background: The Centers for Medicare & Medicaid Services (CMS) uses clinical process-of-care measures in pay-for-reporting, public accountability, and value-based purchasing initiatives to promote better-quality care and improve patient outcomes for Medicare beneficiaries. The research team examined whether improved performance on clinical quality measures is associated with improvements in health outcomes among eligible Medicare beneficiaries.

Methods: The research team applied a structured evaluation process to 171 measures from eight CMS quality measurement programs to identify measures for which a meaningful process-outcome assessment was feasible. The evaluation criteria included the availability of an outcome measure in CMS data to test the linkage between the process and outcomes, the potential size of the expected effect, and the availability of data to adjust for differences in case mix between eligible patients who did and did not receive the process of care. For the selected measures, the research team estimated process-outcome relationships using patient-level data for the Medicare fee-for-service (FFS) population between 2006 and 2012, adjusting for treatment group differences with propensity score methods.

Results: The measure selection process identified the following for evaluation: 10 Hospital Inpatient Quality Reporting Program measures (three heart failure [HF] measures, seven surgical measures) and two Hospital Outpatient Quality Reporting Program measures (one acute myocardial infarction measure and one surgical

measure). The measure screening process excluded 122 measures that were not clinical process measures or for which outcome data were unavailable. Also excluded were 37 process measures that lacked sufficient power to detect a difference. In analyses of more than 3.9 million inpatient and nearly 200,000 outpatient measurement episodes, positive statistically significant process-outcome associations were observed for most of the 12 evaluated measures; however, the clinical significance of these effects was highly variable, and a small number of measures were estimated to have a major impact on the health of the eligible Medicare population during program implementation. Since the beginning of CMS measurement, improved *HF-3: Angiotensin Converting Enzyme Inhibitor (ACEI) or Angiotensin II Receptor Blocker (ARB) for Left Ventricular Systolic Dysfunction (LVSD)* (HF-3) adherence was associated with as many as 5,600 fewer deaths within one year of care episodes, and timeliness of perioperative prophylactic initiation (*SCIP-Inf-1: Prophylactic Antibiotic Received Within One Hour Prior to Surgical Incision* [SCIP-Inf-1]) and discontinuation (*SCIP-Inf-3: Prophylactic Antibiotics Discontinued Within 24 Hours After Surgery End Time* [SCIP-Inf-3]) were each associated with as many as 2,000 and 2,400 fewer post-operative infections, respectively. The remaining evaluated measures were associated with more modest reductions in adverse health outcomes among the Medicare FFS population. Precise estimates of the health benefits of improved process adherence are not possible with existing data because of potential study biases. However, a cumulative estimate of lives saved from the Heart Failure measure set combined (*HF-1: Discharge Instructions*, *HF-2: Evaluation of Left Ventricular Systolic Function*, and *HF-3*), due to increases in process delivery since measurement began, is in the range of 7,000 to 10,000. Additionally, a cumulative estimate of infections averted due to process improvements for the surgical process measure set combined (*SCIP-Inf-1*, *SCIP-Inf-3*, and *SCIP-Inf-9: Urinary catheter removed on Postoperative Day 1 [POD 1] or Postoperative Day 2 [POD 2] With Day of Surgery Being Day Zero*) is in the range of 4,000 to 7,000.

Conclusions: For a limited number of process-of-care measures that could be assessed, the research team found improved adherence to inpatient and outpatient process-of-care measures was associated with decreased mortality and decreased surgical complications in the FFS population during 2006–2012. However, because of the difficulties of controlling for concurrent and unrelated quality improvement efforts, it was not possible to conclude whether the observed relationships are causal. In addition, the absence of a measureable outcome in CMS data and insufficient data to control for differences in case-mix between those who do and do not receive the process of care precluded the ability to assess the impact on health outcomes for a majority of the process measures currently monitored by CMS.

Background

CMS has been using clinical process-of-care measures in pay-for-reporting, public accountability, and value-based purchasing initiatives. These initiatives seek to promote better-quality care and improve patient outcomes for Medicare beneficiaries.^{1, 2} To quantify the relationship between improved delivery of recommended clinical processes of care and patient health outcomes, the research team examined the association between a subset of CMS process-of-care measures and health outcomes of Medicare patients.

A main motivation for evaluating process-outcome relationships in the community is to assess the generalizability and population impact of clinically efficacious processes. Most CMS process measures are supported by clinical trial evidence that has demonstrated improved patient outcomes. Randomized clinical trials have been the gold standard for testing the efficacy of health interventions; however, a trial may be conducted with a narrowly defined subset of the target population, raising questions about the generalizability of the results in real-world settings.³ Assessing the effectiveness of interventions in the community is, therefore, essential to understanding the impact of clinical processes of care. However, unlike randomized clinical trials, observational studies of treatment effectiveness are vulnerable to selection bias, as patients who receive treatment in natural settings might be systematically different from those who do not. For this reason, evaluations of interventions in the community require careful consideration of biases common to observational studies and, where possible, the application of statistical methods to help reduce potential biases.

A number of recent community-based studies (including several systematic reviews) have shown either no effects or mixed results when aiming to link adherence to process measures that are subject to public reporting or pay-for-performance incentives with improved outcomes in the general Medicare population.^{1, 2, 4-10} However, it is uncertain how well these studies addressed four methodological challenges, specifically (1) inadequate power to detect an effect, (2) unmeasured confounding, (3) attribution of differences in individual outcomes to differences in provider-level performance (i.e., ecological bias), and (4) problematic or changing measure specifications over the study period.

For example, in an evaluation of an acute myocardial infarction (AMI) measure, *Beta Blocker Prescribed at Discharge*,^{xcviii} which was one of the first monitored by the CMS Hospital Compare program, Werner and Bradlow found a small, statistically significant improvement of 0.1 percent in one-year mortality between top-performing and bottom-performing hospitals. They concluded that the AMI measure was “not tightly linked to outcomes.” In their analysis, the hospitals in the 75th percentile of performance were 10 percentage points greater than hospitals in the 25th percentile of performance (97 percent versus 87 percent compliance) on the AMI measure. When adherence was nearly complete (i.e., “topped out”) for all hospitals, distinguishing “low” and “high” performance became difficult. However, in patient-level analyses, investigators have had greater power to detect differences in performance because the treated and untreated patients could be directly compared.¹¹

^{xcviii} NQF-endorsed title: *Beta-blocker prescribed at discharge for AMI* (NQF #0160).

Modest associations between hospital performance on Hospital Inpatient Quality Reporting Program (Hospital IQR Program) process measures and patient outcomes have been reported by other authors. Jha et al. found that patients with AMI who were discharged from hospitals in the top quartile of Hospital Quality Alliance (HQA)/Hospital IQR Program performance for AMI process measures had 11 percent lower odds of dying than patients who were discharged from hospitals in the bottom quartile (odds ratio 0.89, 95 percent confidence interval 0.85, 0.94).¹² Similar results were found for patients admitted for HF and pneumonia. Petersen et al. found that better performance on a broader set of AMI measures was associated with lower in-hospital mortality in a small group of hospitals participating in a quality-improvement initiative (odds ratio 0.90, 95 percent confidence interval: 0.84–0.97).¹³

Assessing the implications of these prior studies is difficult because a number of methodological limitations may hinder the validity of the findings. Previous studies of process-outcome associations largely focused on performance at the hospital level. If the primary interest is to determine whether patients who receive a process-of-care have better health outcomes than those that do not, a patient-level analysis is more appropriate than a hospital-level analysis because outcomes in treated and untreated patients can be directly compared.¹¹ Prior process-outcome studies have also been criticized for not adequately controlling for treatment selection bias, which occurs when a patient’s prognosis influences his or her treatment. For example, “healthy candidate bias” is a specific type of selection bias where a patient is less likely to receive a care process because the clinician believes the patient is too ill to benefit.¹⁴ Thus, “healthy candidate bias” in a study can result in findings which incorrectly attribute excessive benefits to process measures. Another methodological issue found in other studies is controlling for hospital fixed effects. Ryan et al. investigated whether the associations observed by Jha et al. and Werner et al. were causal.¹⁵ Ryan et al. included hospital fixed effects to adjust for unobservable characteristics that could affect performance. Ryan et al. found that effect sizes decreased in magnitude and became statistically insignificant after controlling for hospital fixed effects, which suggests that controlling for the inherent quality of a hospital (including general quality improvement practices) may reduce estimates of effect sizes of the relationship between process and outcome.

Also, changes to the specifications of process measures over time have been overlooked in prior research on process-outcome relationships. For example, the longitudinal analyses by Ryan et al., Jha et al., and Werner et al. do not appear to account for changes in the specifications of the process measures over time. Measure specifications refer to definitions for measure denominator (i.e., the population eligible for the process of care, as defined by diagnostic/procedure codes and additional criteria) and numerator (i.e., the patients in the denominator population receiving the process of care). For example, patients with AMI (i.e., denominator population) should be prescribed a beta-blocker at discharge from the hospital (i.e., numerator population), unless they were excluded for various reasons (i.e., denominator exclusions). However, measure specifications are routinely revised relative to measure validity and reliability, which may result in significant changes to the measure population. Although these changes could affect the relationship between process and outcome over time, they have generally not been considered in studies of process-outcome relationships.

Objectives

Two objectives guided the analysis of clinical process-of-care measures and their relationship to patient outcomes and target population outcomes. These objectives are:

1. Determine the association of changes in performance rates for clinical process-of-care measures and changes in patient outcomes between 2006 and 2012.
2. Assess the impact of changes in performance rates for clinical process-of-care measures on the health outcomes of the target Medicare population between 2006 and 2012.

Methods

The research team undertook a multi-stage effort to address methodological problems of prior research on the relationships between process measures and clinical outcomes. First, the feasibility of assessing process-outcome links for currently monitored process-of-care measures was evaluated considering both practical and statistical criteria. For selected measures, the research team used causal inference methods with patient-level data that controlled for a broad array of patient- and provider-level differences to minimize observable differences in the case mix of patients who did and did not receive the process measure (Objective 1). Further, population effects of improved process adherence were estimated by combining the estimated effect of the process with the increase in patients treated with the process during program implementation (Objective 2).

Measures and Data Sources

Process Measure Prioritization

CMS quality measurement programs include hundreds of measures, some of which were not relevant to the study objectives. The research team therefore developed a qualitative and quantitative assessment strategy to identify and prioritize process-of-care measures suitable for inclusion in the analyses.

Qualitative Assessment

The research team reviewed 171 measures from eight CMS measurement programs (Nursing Home Quality Initiative [NHQI]; End-Stage Renal Disease Quality Incentive Program [ESRD QIP]; Medicare Electronic Prescribing Incentive Program [eRx Incentive Program]; Home Health Quality Reporting Program [HH QRP]; Hospital IQR Program; Hospital Outpatient Quality Reporting Program [Hospital OQR Program]; and Medicare Star Rating Programs, Part C and Part D) to determine their relevance to the study objectives and the feasibility of their evaluation using the following criteria. See Appendix i-4 for a hyperlink to a list of measures used in this analysis. The list includes the NQF endorsement status, NQF number if endorsed, and both the measure title used by the CMS program and the measure title used by NQF.

The research team:

1. Excluded outcome and intermediate outcome measures (Criterion #1).
2. For the remaining process-of-care measures, aimed to identify a measureable health outcome targeted by the measure (Criterion #2).

3. Excluded process measures that have not demonstrated a direct influence on clinical outcomes (Criterion #3).
4. Further excluded measures where evidence of potential treatment selection bias was strong and would be difficult to minimize with available data (Criterion #4).
5. Excluded measures whose outcomes could not be observed within a year (Criterion #5) or could not be obtained from administrative claims data (Criterion #6).

Table 8-1 depicts the setting and programs with measures included in this study.

Table 8-1: Programs Initially Included in Analysis

Setting	Program	Abbreviation
 Hospital	Hospital Inpatient Quality Reporting Program	Hospital IQR Program
	Hospital Outpatient Quality Reporting Program	Hospital OQR Program
 Ambulatory	Medicare Electronic Prescribing Incentive Program	eRx Incentive Program
	Medicare Part C (Display or Star Ratings)	Part C
	Medicare Part D (Display or Star Ratings)	Part D
 Post-Acute	Nursing Home Quality Initiative	NHQI
	Home Health Quality Reporting Program	HH QRP
	End-Stage Renal Disease Quality Incentive Program	ESRD QIP

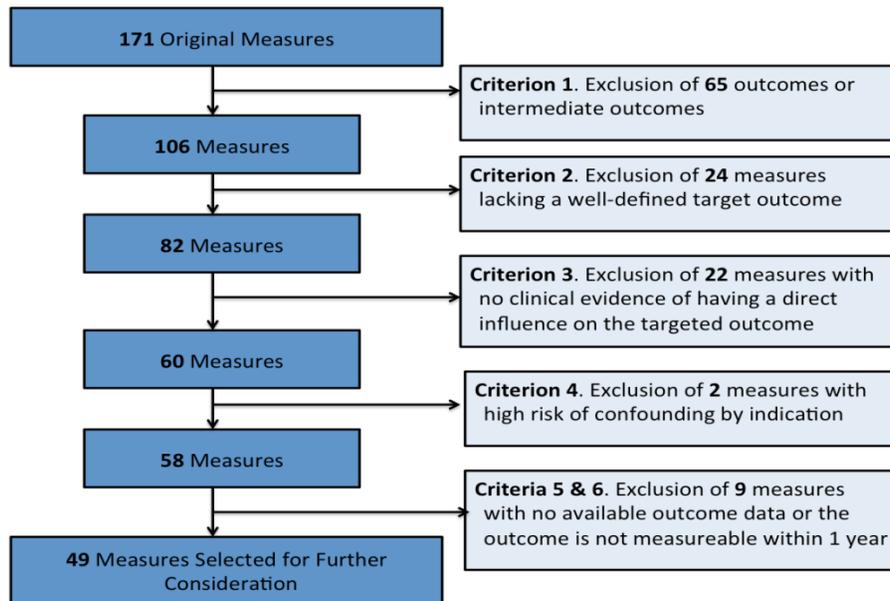
Outcome Prioritization

The research team strived to identify relevant outcomes that can be obtained from existing data for each process measure (e.g., mortality in the Medicare eligibility file). The research team identified relevant outcomes through a scan of the intended health benefits described in the documentation for each measure and supporting studies cited in the National Quality Measures Clearinghouse website. Outcomes that were endpoints reported in at least one published study of a given process measure were considered. Although 30-day outcomes have been the focus of most public reporting, the research team favored longer-term outcomes, as it is more clinically meaningful to demonstrate improved health over a longer period of time if such outcomes were believed to be detectable in available data. For example, most Surgical Care Improvement Project (SCIP) measures were deemed likely to have a stronger effect on short-term outcomes than long-term outcomes. However, the *SCIP-Card-2: Surgery Patients on Beta-Blocker*

Therapy Prior to Arrival Who Received a Beta-Blocker During the Perioperative Period^{xcix} (SCIP-Card-2) measure, which encourages continuing beta-blockers in patients undergoing major surgeries, would be likely to help reduce one-year mortality,^{16, 17} which the research team therefore considered to be a primary outcome.

Three clinical experts independently reviewed the list of identified outcomes to (1) determine the set of outcomes that could be measured from administrative data, and (2) designate the outcome as *primary* or *secondary*. The primary outcome for each analysis was defined as a clinical event that was under the direct influence of the process measure. Although 30-day outcomes have been the focus of most public reporting, the research team favored longer-term outcomes. They are more clinically meaningful to demonstrate improved health over a longer period of time, if such outcomes were detectable in available data. For example, process measures concerning prophylaxis for venous thromboembolism (VTE) would have episodes of VTE treatment noted as primary outcomes. The research team also included mortality as an outcome, but considered it secondary for this measure because prophylaxis against VTE events would not be expected to reduce mortality except by reducing the incidence of VTE. Project-team clinicians resolved disagreement regarding the choice of primary outcome by engaging in a group discussion that carefully considered measure specification documents and relevant published literature. Subsequently, this qualitative assessment and outcome prioritization process identified 49 process measures and associated outcomes that met the study eligibility criteria, as shown in Figure 8-1.¹⁸

Figure 8-1: Summary of Qualitative Assessment



^{xcix} NQF-endorsed title: *Surgery Patients on Beta-Blocker Therapy Prior to Arrival Who Received a Beta-Blocker During the Perioperative Period* (NQF #0284).

Quantitative Assessment

Because it would not be feasible to conduct thorough analyses of association for all measures that met the study’s qualitative criteria, measures were further prioritized by their expected Type I error (the probability of falsely concluding a process-outcome exists when it does not) and power (the probability of detecting a true process-outcome association). The research team based the calculations of error and power on publicly available data on performance rates and measure denominators from Hospital Compare, Nursing Home Compare, and Home Health Compare. Expected effect size information was obtained from the highest-quality evidence available for the target population of the CMS measure and was the basis for the power calculation. The error and power assessment accounted for several important features of patient-level analyses of CMS measurement programs, including omitted variable bias, clustering of patients within providers, and high process rates. Under these conditions, measures with a Type I error closest to the target 5 percent level and greatest power to detect the expected effect size were given higher priority in the study analysis. Table 8-2 shows the programs with measures that were prioritized for this analysis and Table 8-3 displays the associated measures and primary outcomes.

Table 8-2: Programs Included in Analysis

Setting	Program	Abbreviation
 Hospital	Hospital Inpatient Quality Reporting Program	Hospital IQR Program
	Hospital Outpatient Quality Reporting Program	Hospital OQR Program

Table 8-3: Process Measures and Primary Outcomes Prioritized for the Study^c

Measure	Primary Outcomes
Hospital Inpatient Quality Reporting Program	
<i>HF-1: Discharge Instructions</i>	All-cause mortality within one year Congestive heart failure (CHF) readmission or all-cause mortality within one year
<i>HF-2: Evaluation of Left Ventricular Systolic (LVS) Function</i>	All-cause mortality within one year CHF readmission or all-cause mortality within one year
<i>HF-3: Angiotensin Converting Enzyme Inhibitor (ACEI) or Angiotensin Receptor Blocker (ARB) for Left Ventricular Systolic Dysfunction (LVSD)</i>	All-cause mortality within one year CHF readmission or all-cause mortality within one year
<i>SCIP-Inf-1: Prophylactic Antibiotic Received Within One Hour Prior to Surgical Incision</i>	Secondary diagnosis of wound infection given as discharge diagnosis at the end of index admission not present on admission (POA) or death OR readmission for wound infection within 30 days

^c See Appendix i-4 for a hyperlink to a crosswalk of all measures in Table 8-3, including CMS Title, NQF Title, NQF I.D., and NQF endorsement status. The data used for this analysis are retrospective. See Table 8-7 for further information.

Table 8-3: Process Measures and Primary Outcomes Prioritized for the Study^c

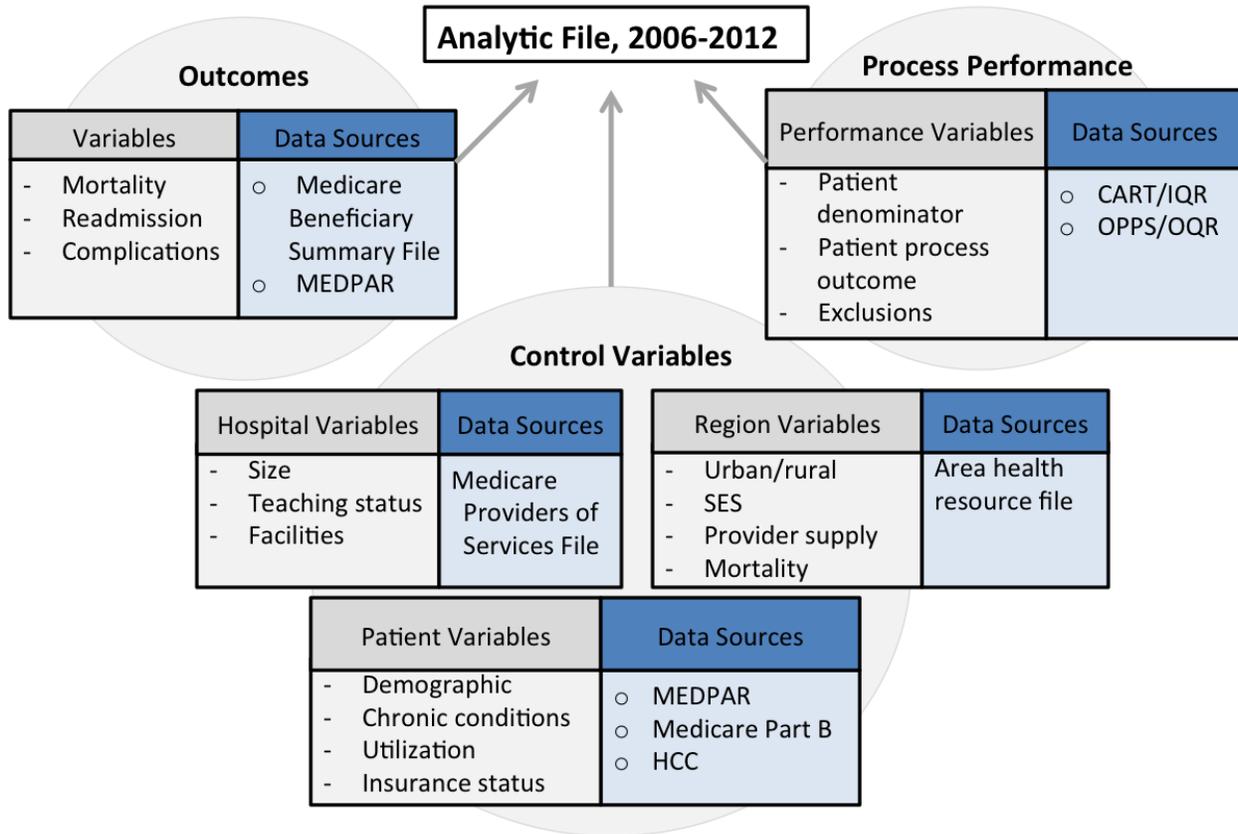
Measure	Primary Outcomes
<i>SCIP-Inf-3: Prophylactic Antibiotics Discontinued Within 24 Hours after Surgery End Time</i>	Secondary diagnosis of <i>C. difficile</i> infection not POA or death or readmission within 30 days of discharge for <i>C. difficile</i> infection
<i>SCIP-Inf-6: Surgery Patients with Appropriate Hair Removal</i>	Secondary diagnosis of wound infection not POA or death or readmission within 30 days of discharge for wound infection
<i>SCIP-Inf-9: Urinary Catheter Removed on Postoperative Day 1 (POD 1) or Postoperative Day 2 (POD 2) With Day of Surgery Being Day Zero</i>	Secondary diagnosis of urinary tract infection (UTI) not POA <i>any time</i> during index admission or death or readmission within 30 days of discharge for UTI
<i>SCIP-Card-2: Surgery Patients on Beta-Blocker Therapy Prior to Arrival Who Received a Beta-Blocker During the Perioperative Period</i>	All-cause mortality within one year
	CHF readmission or all-cause mortality within one year
<i>SCIP-VTE-1: Surgery Patients With Recommended Venous Thromboembolism Prophylaxis Ordered</i>	Secondary diagnosis of deep vein thrombosis (DVT) or pulmonary embolism (PE) not POA or death or readmission within 30 days of discharge for DVT or PE
<i>SCIP-VTE-2: Surgery Patients Who Received Appropriate Venous Thromboembolism Prophylaxis Within 24 Hours Prior to Surgery to 24 Hours After Surgery</i>	Secondary diagnosis of DVT or PE not POA or death or readmission within 30 days of discharge for DVT or PE
Hospital Outpatient Quality Reporting Program	
<i>OP-4: Acute Myocardial Infarction (AMI) Aspirin at Arrival</i>	All-cause mortality within 30 days
<i>OP-6: Perioperative Timing of Antibiotic Prophylaxis^{ci}</i>	Admission for wound infection or death within 30 days.

Analytic File Construction

Data were combined from numerous sources to create an analytic file that included patient-level performance data for clinical process measures, healthcare claims data, enrollment data, geographic data, and provider characteristics, from 2006–2012.¹⁸ A dataset was created for each process measure that included all fee-for-service (FFS) beneficiaries with recorded data regarding the process-of-care measure (1=received measure, 0=did not receive measure) and one or more clinical outcomes (1=experienced outcome, 0=did not experience outcome). Each dataset varied in size based on the denominator population but generally included more than 100,000 measurement episodes. Demographic, socioeconomic, facility, and patient-level clinical characteristics were included, which could potentially bias the analysis if omitted. A schematic of the process is depicted in Figure 8-2 and detailed in the sections that follow.

^{ci} The data used for this analysis are retrospective. OP-6 has been removed from the Hospital OQR Program.

Figure 8-2: Diagram of Analytic File Construction



Study Population

The analysis was limited to the FFS population because non-FFS patients lacked complete data on outcomes. For patients who switched from FFS to another type of insurance, only the time periods when the patients were enrolled in FFS were included in the analysis.

Process Measures

The research team obtained data for hospital-based process measures from the Quality Improvement Organization clinical warehouse. The analysis used Hospital IQR Program data collected during 2006–2012 for the inpatient setting, and Hospital OQR Program data collected during 2008–2012 for the outpatient setting. The source data for Hospital IQR Program and Hospital OQR Program provided the principal diagnosis or procedure codes used to determine patient eligibility for the measures as well as reasons for exclusion from the measure denominator.

Outcome Identification

Data on hospitalizations were identified using inpatient claims data from the Medicare Provider Analysis and Review (MEDPAR) files for 2006–2012 and death dates from the Medicare Master Beneficiary Summary File for the same time period. Outcomes included mortality, adverse events noted in the inpatient claim for surgical procedures (such as VTE noted on an admission for a surgery qualified for SCIP measures), and future hospitalizations for a number of conditions (Table 8-3). Details of the codes used to identify these clinical events are provided in a supplementary technical report.¹⁸

Demographic and Socioeconomic Factors

The research team used data from the Master Beneficiary Summary File to determine patient-level demographic status (age, sex, race); income (median income for the enrollee's ZIP code derived from the American Community Survey); enrollment status (FFS or Medicare Advantage); and location.

Clinical Risk-Adjustment Factors

Patient-level risk-adjustment factors were extracted from three sources for all Medicare patients as follows: To avoid reverse causation, only factors present before or at the time of the care episode were considered. For inpatient measurement episodes, the research team used all primary and secondary diagnoses in MEDPAR claims data to compute conditions in the Elixhauser risk algorithm for all admissions.¹⁹ To increase flexibility, the research team used the individual indicators rather than an overall summary score.²⁰ The research team identified the Hierarchical Condition Category (HCC) score from annual CMS-HCC Risk Adjustment data files.^{21, 22} Annual files from the CMS Chronic Conditions Data Warehouse (CCW) database were used to add indicator variables for each of 30 chronic conditions. Although the Elixhauser, CCW, and HCC include some of the same conditions, they referred to different periods of time and were therefore complementary rather than overlapping. Each source for risk adjustment had limitations: Elixhauser indicators were only available for hospital-based measurement episodes, HCC scores were derived from an assessment at the beginning of the calendar year for each patient, and CCW indicators were current as of the most recent assessment within the past 12 months and were missing for patients without pre-existing chronic conditions.

Facility and Regional Characteristics

The research team used the CMS Provider of Services file to generate facility-level variables, including the county code; facility type (e.g., state, local, non-profit, federal); indicator for teaching hospital (if applicable); number of beds; and region (Northeast, South, Midwest, and West). The research team flagged hospitals for important characteristics, including having a cardiac catheterization laboratory or facilities for performing coronary artery bypass graft (CABG) surgeries and linked the Health Resources and Services Administration Area Health Resources File by county to add market-level variables, including the provider-to-population ratio.

Final Analytic File Construction

Final analytic files were constructed beginning with the measurement episode and adding data on demographic, geographic, socioeconomic, and provider characteristics, and patient-level clinical characteristics from the datasets described above. The research team identified outcomes in four ways: First, deaths were identified by flagging patients who died within a set time period after a measurement episode. Second (for surgical process measures), surgical complications were identified during the hospitalization in which the episode took place (if those complications were not flagged as “present on admission”). Third, other outcomes were identified by examining all hospitalizations occurring after the initial measurement date with an outcome listed in Table 8-3. Fourth, the research team constructed composite outcomes in certain analyses (e.g., surgical complications identified within the initial or subsequent hospitalizations during a certain time frame).

Analysis

Objective 1—Determine the Association Between Changes in Performance Rates for Clinical Process-of-Care Measures and Changes in Patient Outcomes

For each prioritized process-outcome pair (Table 8-3), the research team used counts and rates to describe national trends in the denominator, process rate, and outcome rate for the FFS population over time. To evaluate the effect of processes on health outcomes, propensity score methods were used to control for differences between treatment groups. Propensity score methods are a statistical approach that aims to analytically balance treatment groups in the absence of randomization.^{23, 24}

Adjusting for Treatment Group Differences

In real-world settings, eligible patients who receive treatment might be systematically different from those who do not. Because these differences might also influence health outcomes (for example, if patients with a better prognosis were more likely to be treated than patients with a poorer prognosis), the research team used propensity score methods to balance treatment groups on known characteristics that could influence the chance of receiving a process of care or experiencing an adverse health outcome.

Propensity score methods are a popular approach to balance comparison groups in the absence of randomization. The *propensity score* itself refers to the probability that a patient will receive treatment and is typically estimated using regression techniques. Weights based on the inverse of the propensity score can then be used to create a pseudo-randomized comparison, in a similar way that survey weights are used to obtain a pseudo-representation of the survey population. For example, if the observed treatment group had three times more women than the control group, propensity-score weighting would assign higher weights to women in the control group, so that the representation of women in the weighted comparison groups was 1:1.

The research team considered a large number of candidate covariates in the balancing of treatment groups. Candidate covariates included patient demographics; HCC score; 26 chronic conditions recorded in the CCW; and, for inpatient measures, 28 Elixhauser conditions at the time of admission. For patients receiving measured surgical processes of care, the research team

also included procedure type as a covariate to adjust for risk differences associated with major procedure categories; procedural categories were defined using the same classification used for the seven procedural strata in the 2012 sampling plan for SCIP measure reporting.¹⁸ Additional region- and facility-level variables included census division, urbanicity,^{cii} income, mortality rate, and provider-to-patient ratio (for region); and teaching status, bed count, cardiac unit, CABG facility, and catheterization laboratory (for facility). The research team used an approach for selecting covariates that errs on the side of inclusion, as balance was the primary goal and overfitting was a less relevant concern. Covariates having a correlation with process receipt or primary outcome at least 0.01 in magnitude were included in the propensity score model. Effects for age, gender, and race and all of their possible interactions were included in all models.

Propensity score estimates were obtained using the covariate balancing method of Imai and Ratkovic (2014). Standardized mean differences (SMD) and variance ratios were used to diagnose the balance of each covariate before and after propensity score weighting.²⁵ Imbalance was defined as an SMD greater than 0.2 in absolute value or, for continuous variables, a variance ratio less than 0.5 or greater than 2.²⁶

Process Effect Estimation

Adjusted estimates of process-outcome relationships were calculated using propensity-weighted regression. For mortality and infection outcomes, the measure of effect was the absolute risk difference, while for readmission outcomes the effect was the hazard ratio—a measure of relative risk for a time-to-event variable—and death was treated as a censoring event. Covariates with more than minimal imbalance after propensity-score adjustment (defined as an adjusted SMD of 0.05 or more) were included in the outcome regression model. This approach is known as *doubly-robust* estimation because the use of both a propensity-score weight *and* regression adjustment provides two opportunities to control for differences between treatment groups.²⁷

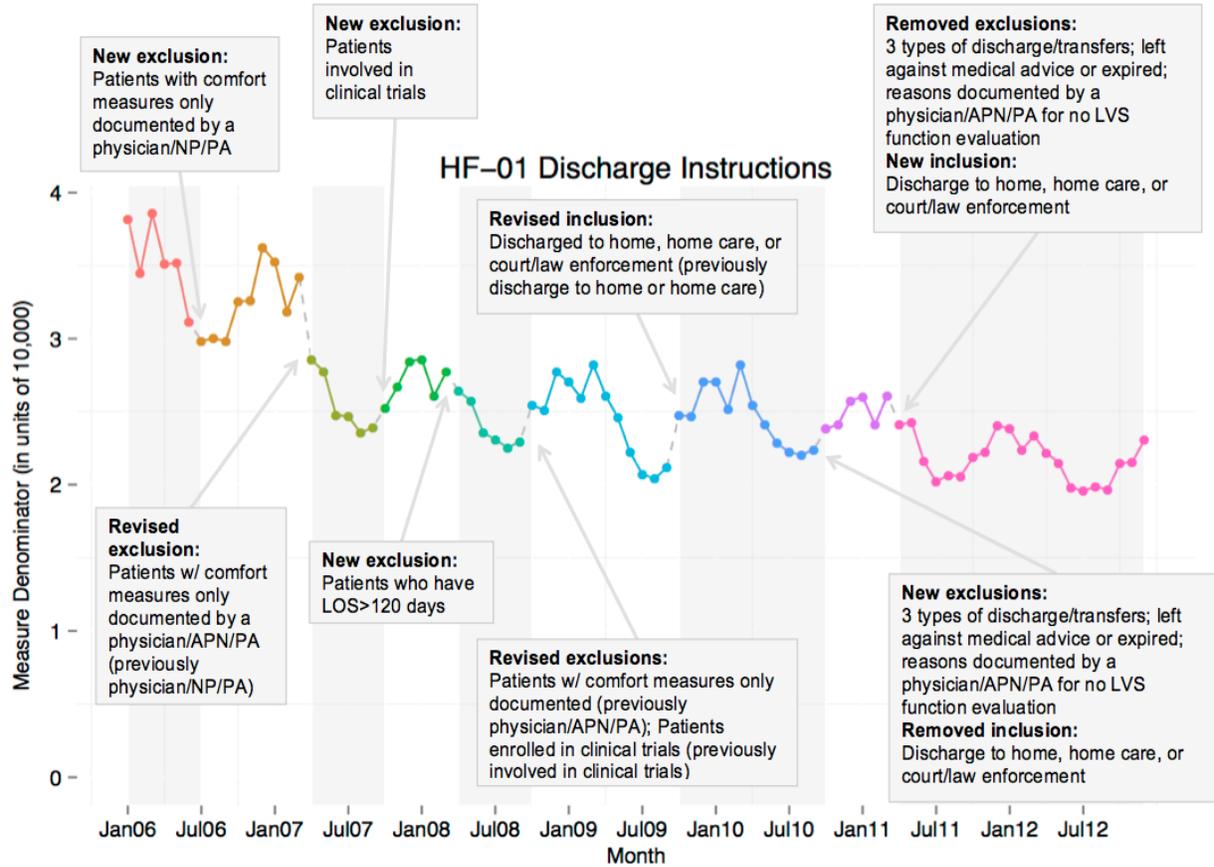
Specification Changes

The present analyses include multiple years of process delivery and outcomes. Over this time, the specifications defining the inclusion and exclusion criteria for each process measure have changed. These changes could result in a different case mix of patients from one specification period to the next. Of particular concern are revisions that result in significant changes to the size and risk level of the measure population. Figure 8-3 provides an illustrative timeline of one set of changes for the Hospital IQR Program measure *HF-1: Discharge Instructions*^{ciii} (HF-1). Between 2006 and 2012, HF-1 underwent eight changes to the definition of its measure population (a number typical for Hospital IQR Program measures). For example, in April 2008, the HF-1 measure specification was changed to exclude patients with lengths of stay of 120 days or greater. The research team expected that this change would improve the prognostic status of the entire HF-1 population.

^{cii} Urbanicity refers to the degree to which a geographical unit is urban.

^{ciii} Previously NQF-endorsed title: *Heart Failure (HF): Detailed discharge instructions* (NQF #0136).

Figure 8-3: HF-1 Denominator Population versus Measure Specifications in 2006–2012



To identify unusual changes in population definitions, the research team documented the changes with each specification version. The research team compared the monthly change in denominator size and mortality rate for periods with and without a specification change. Specifications associated with a comparatively large change in denominator percentage and mortality rate were flagged as changes with a potentially high impact on patient heterogeneity.

Owing to possible between-period differences in the measure population, the research team separated each process-measure effect analysis by specification period. The research team then evaluated differences in effects by period using meta-analytic techniques. In other words, the result within each specification period was treated as a separate study, and the combined effect was estimated in similar fashion to a meta-analysis. The research team evaluated between-period heterogeneity with the Q-test and Higgins' I^2 ; ²⁸ estimates of the overall effect of processes across periods were derived using a meta-analytic random-effects model, where the specification period was the group variable rather than study.

Secondary Analyses

In addition to looking at overall treatment effects, it is important to examine variation in process effects by hospital. Large between-hospital variation in process effects would indicate unexplained differences in process-outcome links, possibly resulting from hospital or case-mix

factors not included in the analysis. To estimate between-hospital variation, the research team expanded the fully adjusted regression model of process effects to include hospital random effects for the outcome base rate and process effect.

The study's main analyses focus on the average treatment effect (ATE), which is the expected effect of treatment for the complete measure population. There may also be interest in the expected effect among patients who did *not* receive the process, the so-called average treatment effect among controls (ATC). The ATC indicates the change in outcomes to be expected if process rates were to increase above current levels. For each prioritized process-outcome pair, the research team estimated the ATC using the 1:2 matched sample described below.

Finally, because clinical studies have evaluated shorter-term outcomes for HF-1, a secondary analysis was conducted to assess the association of this process with 30-day mortality or CHF readmission.

Sensitivity Analyses

Observational studies in which treatment effects are not experimentally assigned may have several biases. For example, confounding by indication, where providers may choose not to provide a process of care because of a patient's prognosis, is a major concern. To protect against this confounding, the research team controlled for a large number of patient variables. Unfortunately, data for a number of well-known risks, such as smoking, current medications, or individual socio-economic status, were not available. Therefore, the research team conducted a number of analyses to assess the threat of the findings to confounding biases.

To assess the potential risk of bias due to treatment group imbalance, the research team contrasted the doubly-robust risk differences to unadjusted risk differences of process effect. The research team also performed several analyses to evaluate the sensitivity of the main findings to the doubly-robust propensity score methodology. One set of sensitivity analyses focused on the ability to balance the treated and control groups using matching. Under matching, each control subject is matched to the two treated subjects with the most similar set of covariates (1:2 matched sample), rather than propensity score weighting. Given that many more patients received treatment than did not receive treatment, 1:2 matching yielded a sample that was representative of the control patients but not necessarily representative of treated patients.

A second set of sensitivity analyses contrasted the doubly-robust effect estimates to estimates with propensity score weighting and estimates using regression adjustment for all selected covariates without propensity score weighting. The research team also calculated the c-statistic for each process measure-outcome analysis and period. The c-statistic measures the model's discriminatory ability; it ranges from 0.5 to 1.0, with 1.0 indicating the highest possible discrimination. Better discrimination indicates that differences in risk are well-explained by the set of risk adjusters included in the model, and there may be less risk that an important confounder has been omitted.

All methods of adjustment for observable imbalances between treatment groups are limited by characteristics the research team was able to *observe*. Even the best adjustment method could

produce biased findings if there is unmeasured confounding (i.e., there are important contributors to process, and health outcomes are not included in the adjustment).

The research team used two approaches to assess the impact of unmeasured confounding on the main findings. First, the set of covariates was expanded in the adjusted analyses to incorporate a number of measures derived from the CMS-Part B physician file.¹⁸ The additional measures included Elixhauser conditions, total visits, total specialty visits, total major procedures, and use of erythropoietin in the six months prior to the encounter episode. These factors were included to adjust for differences in the severity of kidney disease between treatment and non-treatment groups. Second, because there were potentially significant clinical variables that could not be captured with claims data, the research team conducted analyses to determine the sensitivity of the findings to a hypothetical omitted variable. The impact of an omitted confounder is driven by the difference of the prevalence of the confounder between treatment groups and the magnitude of the association of the confounder with the health outcome after adjusting for observed variables.²⁹ The research team examined the maximum prevalence differences and outcome effects among the observed variables to determine the most extreme yet plausible values for the omitted variable. The team then made the simplifying assumption of equal prevalence differences and outcome effects of the unmeasured confounder across the strata defined by observed covariates.

Objective 2—Assess the Impact of Changes in Performance Rates for Clinical Process-of-Care Measures on the Health Outcomes of the Targeted Medicare Population

To assess the impact of increased process-of-care delivery for the targeted Medicare population, the research team estimated the total number of adverse clinical outcomes that were possibly averted with increased rates of adherence between 2006 and 2012. The research team first estimated the number of newly treated beneficiaries in each month since CMS began to monitor the measure. The estimate of newly treated beneficiaries in each month was equal to the patient denominator in that month multiplied by the difference in the month's process rate and the process rate at the beginning of program measurement. Calculating the newly treated beneficiaries in each month accounts for seasonal trends and other sources of month-to-month variation that may influence the total size of the measure population. With the estimate of the additional treated patients, the total number of averted events was obtained by multiplying the number of newly treated patients by the estimated overall absolute risk reduction attributable to the process measure.

Additional steps were taken to avoid double-counting averted events for measure populations with a large overlap in denominators. The number of newly treated *HF-3: Angiotensin Converting Enzyme Inhibitor (ACEI) or Angiotensin Receptor Blocker (ARB) for Left Ventricular Systolic Dysfunction (LVSD)*^{civ} (HF-3) patients was subtracted from the *HF-2: Evaluation of Left Ventricular Systolic (LVS) Function*^{cv} (HF-2) population, which conservatively assumes that all of the benefit for patients who receive both HF-2 and HF-3 is due to HF-3. Similarly, because the overlap for the *SCIP-VTE-1: Surgery Patients With Recommended*

^{civ} NQF-endorsed title: *ACEI or ARB for left ventricular systolic dysfunction—Heart Failure (HF) Patients* (NQF #0162).

^{cv} NQF-endorsed title: *Evaluation of Left ventricular systolic function (LVS)* (NQF-0135).

Venous Thromboembolism Prophylaxis Ordered^{cvi} (SCIP-VTE-1) and *SCIP-VTE-2: Surgery Patients Who Received Appropriate Venous Thromboembolism Prophylaxis Within 24 Hours Prior to Surgery to 24 Hours After Surgery*^{cvi} (SCIP-VTE-2) populations is nearly 100 percent, the team conservatively assumed that the newly treated beneficiaries for SCIP-VTE-2 represented the total number of patients who could benefit from SCIP-VTE-1 and SCIP-VTE-2.

Limitations

While addressing shortcomings of prior work, the present study has important limitations. To have sufficient information about outcomes and patient medical histories to conduct adjusted patient-level analyses, the research team had to limit the study to the Medicare population aged 65 years and older—for which more information is available—and could not determine whether the associations found also apply to younger populations. Despite the extensive administrative health data collected about the Medicare population, the research team lacked many important clinical details that could inform beneficiaries' prognostic status at the time of presentation (e.g., medications, functional status, and mental status). Furthermore, many post-surgical outcomes (including UTI and VTE) are frequently miscoded, as are other hospitalization outcomes.³⁰ Thus, the research team cannot rule out bias due to measurement error or unmeasured confounding.

Clinicians may choose to provide treatment based on a patient's prognosis, which leaves analyses of associations of process measures with outcomes vulnerable to significant confounding bias. The research team found several reasons to suspect that the risk of this type of bias for the process-of-care measures evaluated was high. Process adherence rates have been steadily increasing for all of the measures and have topped out for a smaller number of measures. Given the incentives to achieve high rates of adherence, clinicians might choose not to give a care process when there are clinically justifiable reasons not to do so, such as a treatment contraindication. The imbalances observed between the patients who received and did not receive process measures, such as the disparity in the prevalence of renal failure, provide strong evidence of treatment by indication. Although the research team was able to make balanced comparisons for observed patient differences, it is likely that important confounders were omitted. Based on the results of sensitivity analyses that assessed the potential threat of omitted confounder bias, the research team concluded that the reported positive effects of processes may overestimate the true benefits of these processes.

This study focuses on a subset of process measures because many currently monitored measures lack a well-defined targeted outcome or sufficient data about the health outcomes targeted. While the research team could identify outcomes evaluated in the clinical trial or other clinical studies supporting the recommended process measures, these outcomes were frequently not available or not easily constructed using administrative data. When an available outcome measure was not an exact match for what was evaluated in the clinical trial, judgment was required about whether the available outcome measure was a reasonable proxy. Because proxies

^{cvi} Previously NQF-endorsed title: *Surgery Patients with Recommended Venous Thromboembolism (VTE) Prophylaxis Ordered* (NQF #0217).

^{cvi} NQF-endorsed title: *Surgery Patients Who Received Appropriate Venous Thromboembolism Prophylaxis Within 24 Hours Prior to Surgery to 24 Hours After Surgery* (NQF #0218).

are imperfect substitutes for the targeted outcomes, caution is needed when interpreting process-outcome associations using a proxy outcome.

Furthermore, for process-outcome pairs that could be analyzed with present administrative data sources, several features of measure programs complicate analyses of associations. For most measures, hospitals with a sufficiently large number of eligible patients can report on performance for a sample rather than the complete measure population. If process effects differ for hospitals that sample, the failure to adjust for sampling rates could result in biased process-outcome estimates.

Also, as described in this report, the measure populations for each process measure frequently changed over time due to revisions of measure specifications. While these revisions are intended to improve the targeting of the process measure, they could complicate process-outcome analyses if patients in one specification period are not comparable to patients in another specification period. The research team closely documented specification changes for the 12 measures presented in this report and found evidence of between-period heterogeneity for most measures. This indicates that summary estimates of process effects that combine results across specification periods should be interpreted with caution. One potential way of addressing the measure specification changes is to have the measure developers quantify the magnitude of population change whenever a specification change affects the population of interest.

There were two important additional error sources in the assessment of the population impact of increased process adherence—sampling, and overlapping measure populations. Sampling refers to a hospital’s ability to report process rates for a sample of eligible patients, if the population of eligible patients meets a designated population size. The research team lacked access to reliable information on sampling fractions that would have allowed them to correct the estimates to account for sampling. In the absence of bias in the process effect estimate, the presence of sampling would tend to result in an underestimate of the actual population benefit. At the same time, for simplicity, the research team ignored overlap in measure populations, which could result in overstating the total expected events averted. Because the relative magnitude of these counteracting inaccuracies are not known, the research team could not determine whether the estimated population impact is more likely to be under- or overestimated.

Several of the measures evaluated in this study have subsequently been suspended or removed from the Hospital IQR Program,³¹ including HF-1³², HF-3³², *SCIP-Inf-6: Surgery Patients with Appropriate Hair Removal*^{cviii} (SCIP-Inf-6)³³, and SCIP-VTE-1.³⁴ The present status of process measures was not a consideration in the study’s measure selection criteria because the objective of the work was to evaluate the historical impact of processes monitored by CMS between 2006 and 2012. However, it is notable that three of the four removed measures were not shown to have had a meaningful impact on population health in the present study. This neither validates nor invalidates the removal of these measures because effectiveness is frequently one part of the criteria that go into this decision. However, it does provide support for concerns about the continued value of monitoring these measures.

^{cviii} NQF-endorsed title: *Surgery patients with appropriate hair removal* (NQF #0301).

Finally, the estimated population health improvements represent a combination of causes that include the effect of general increases in quality improvement in addition to the direct effect of CMS national incentive programs. However, the research team could not distinguish among these potential causes. An evaluation of the specific impact of CMS programs on patient health outcomes would require detailed data on measure process adherence, clinical outcomes, and other clinical data for a period before and after program implementation. In addition, the evaluation would ideally be able to assess a control group in which the program was not implemented. However, adding a control group and requiring additional data collection may not be possible in a nationally implemented program.

Results

Measure Selection

Of the 171 measures screened for inclusion, the research team excluded:

- ◆ Sixty-five measures that were outcomes (e.g., Hospital IQR Program: Pressure ulcers).
- ◆ Twenty-two for which a targeted outcome could not be identified (e.g., Hospital IQR Program: Participation in a systematic database for stroke care).
- ◆ Twenty-four lacking evidence of a clinical effect on their targeted outcome (e.g., Hospital IQR Program: *SCIP-Inf-10: Surgery Patients With Perioperative Temperature Management*^{cix}).
- ◆ Two with a high risk of confounding by indication (e.g., Nursing Home Minimum Data Set 3.0: “Percent of residents who have/had a catheter inserted and left in their bladder (long stay).”)
- ◆ Nine with a targeted outcome that could not be obtained from available data (e.g., Medicare Part C: Diabetes care—kidney disease monitoring) or could not be observed within one year (Figure 8-1).

When a power assessment was applied to prioritize the 49 measures that passed the qualitative review, no process-outcome pair had adequate power in the presence of a probable confounding bias.¹⁸ Since it was not possible to limit the analyses to measures with a level of power typically required of experimental studies, the research team selected 12 measures with the greatest power and lowest inflation of Type I error among the group of 49 measures. These measures included three HF and seven SCIP measures from the Hospital IQR Program and two measures from the Hospital OQR Program. Table 8-3 provides a full description of the selected measures and their primary outcomes.

Denominator and Process Rate Trends

From 2006 to 2012, measure denominators (i.e., the eligible patient population for the clinical process) showed substantial month-to-month variation.¹⁸ Possible explanations for the instability in measure population include secular trends, seasonal variation, or specification changes. The research team found that specification changes for the HF-1, *SCIP-Inf-1: Prophylactic Antibiotic Received Within One Hour Prior to Surgical Incision* (NQF #0527) (SCIP-Inf-1),

^{cix} Previously NQF-endorsed title: *Surgery Patients with Perioperative Temperature Management* (NQF #0452).

SCIP-Inf-3: Prophylactic Antibiotics Discontinued Within 24 Hours after Surgery End Time (NQF #0529) (SCIP-Inf-3), SCIP-VTE-1, and SCIP-VTE-2 measures were frequently followed by large changes in the measure denominator population size and mortality rate.

Despite the volatility in the measures denominator populations (i.e., the counts of beneficiaries eligible for the measures), the percentage of eligible beneficiaries who actually received the recommended process of care consistently increased over time for all measures considered in this report.¹⁸ The yearly rate of increase ranged from 1 percent to 4 percent across the prioritized measures, with HF-1 and SCIP-Inf-3 showing the highest rates of increase. All Hospital IQR Program measures had a national process rate above 95 percent, except HF-1, which was 93 percent. These trends may reflect general improvements in process rates for the FFS population or changes in the inclusion and exclusion criteria for process measures.

Patient Population

Patient-level analyses included episodes of care for 1.3 million Medicare patients eligible for the HF measures (HF-1, HF-2, and HF-3); 2.6 million eligible for at least one of the SCIP-Inf, SCIP-Card-2, and SCIP-VTE measures; and more than 170,000 outpatient care episodes for at least one of the *OP-4: AMI Aspirin at Arrival* (NQF #0286) (OP-4) and *OP-6: Perioperative Timing of Antibiotic Prophylaxis*^{cx} (OP-6) measures (Table 8-4). Eligible patients were typically 70 to 75 years of age and had multiple pre-existing conditions, with patients eligible for the HF measures being on average sicker than patients in the other measure populations.

Table 8-4: Percentage of Measure Population for Grouped Measures and the Current Specification Period^{cx}

Covariate	HF ^{cxii}	SCIP-Inf ^{cxiii}	SCIP-Card-2	SCIP-VTE ^{cxiv}	OP-4	OP-6
Total Population (millions)	1.3	1.6	0.2	0.8	0.07	0.01
Age, Mean (interquartile range)	75.2 (68,84)	72.2 (67,78)	73.1 (68,79)	72.1 (67,79)	70.6 (65,79)	72.2 (67,81)
Race						
White, Non-Hispanic	77.3	88.8	88.7	88.2	89.4	86.4
Black, Non-Hispanic	17.9	7.4	7.7	7.9	7.3	9.4
Hispanic	2.1	1.2	1.0	1.3	1.0	1.5
Asian	0.9	0.7	0.7	0.7	0.4	0.8
Other ^{cxv}	1.7	1.7	1.6	1.7	1.7	1.7
Dual Eligible	22.1	12.7	12.6	13.9	21.8	16.1
Median income of zip (\$10,000)	4.8	5.0	4.9	5.0	4.3	4.9
ESRD	6.0	1.0	2.0	1.0	3.0	2.0

^{cx} Previously NQF-endorsed title: *Perioperative Care: Timing of Prophylactic Parenteral Antibiotics—Ordering Physician* (NQF #0270)

^{cx} Dates of most recent specification period of each measure.

^{cxii} All persons in either the HF-1, HF-2, or HF-3 measure population in the most recent specification period.

^{cxiii} All persons in either the SCIP-Inf-1, SCIP-Inf-3, SCIP-Inf-6, or SCIP-Inf-9 measure population in the most recent specification period.

^{cxiv} All persons in either the SCIP-VTE-1 or SCIP-VTE-2 measure population in the most recent specification period.

^{cxv} “Other” includes Asian, North American Native, and Other/unknown racial categories.

Table 8-4: Percentage of Measure Population for Grouped Measures and the Current Specification Period^{cx1}

Covariate	HF ^{cxii}	SCIP-Inf ^{cxiii}	SCIP-Card-2	SCIP-VTE ^{cxiv}	OP-4	OP-6
Hierarchical Chronic Conditions, Mean (interquartile range)	2.4 (1.1,3.3)	1.1 (0.5,1.3)	1.3 (0.6,1.6)	1.1 (0.5, 1.3)	1.5 (0.6,1.9)	1.5 (0.7,1.9)
Chronic Conditions Warehouse (Total), Mean (interquartile range)	6.4 (4,9)	3.7 (1,6)	4.4 (2,6)	3.7 (1,5)	4.5 (2,7)	4.6 (2,7)
Elixhauser Index, Mean (interquartile range)	5.4 (1,8)	2.8 (0,5)	4.1 (0,7)	3.0 (0,5)	N/A	N/A
Teaching Hospital	20.2	19.3	21.0	17.5	3.6	22.9
Hospital Had CABG facility	58.0	59.0	62.5	53.3	10.4	69.6
Hospital Had Catheterization Laboratory	47.0	48.3	50.5	46.1	17.1	53.2
Hospital Had Cardiac Care Unit	79.2	78.0	80.3	75.6	48.7	82.4

Outcome Rates

In 2012, one-year patient mortality rates ranged from 26 percent to 31 percent for patients receiving HF process measures.¹⁸ For the SCIP measure populations that received the process interventions, the 30-day complication rates were 1 percent to 4 percent, while one-year mortality for the SCIP-Card-2 process-of-care population was 6 percent. For the AMI outpatient measure OP-4, the 30-day mortality rate of those receiving the process of care was 5 percent, and the OP-6 process-of-care population had a 30-day wound infection rate of 1 percent. The highest event rates were observed for the HF populations.

Specification Changes

In examining how the measure populations changed with the implementation of new specifications, the research team generally noted an inverse relationship between change in population size and mortality in the measure population; that is, as one increased, the other decreased. A small number of specification changes had a significant impact on the size of the measure population and its mortality rates;¹⁸ changes in the 99th percentile were defined as instances of a “major change” in specifications. Based on this definition, the research team identified four major specification changes:

- ◆ October 2009 specification for the HF-1 measure, which added patients discharged or transferred to a court or law enforcement.³⁵
- ◆ April 2007 specification for the SCIP-Card-2 measure, which excluded patients whose ICD-9-CM principal procedure was performed entirely by laparoscope and added 75 new procedural exclusions.³⁶
- ◆ October 2008 change in exclusion based on procedure length from <30 minutes to less than one hour and change the exclusion of post-operative stay from more than 24 hours to more than three days for the SCIP-VTE-1 and SCIP-VTE-2 measure populations.³⁷
- ◆ April 2011 change in the exclusion of post-operative stay from greater than three days to greater than two nights for the SCIP-VTE-1 and SCIP-VTE-2 measure populations.³⁸

Adjustment for Imbalance

Patients who received a measured process of care differed in important ways from those who did not. Figure 8-4 summarizes the differences in covariates found in the most recent specification period. Each column of the grid corresponds to a process measure and each row a characteristic. Shaded squares denote an imbalance for a given characteristic, where imbalance is measured by the standardized mean difference (SMD). Negative imbalances (denoted in red) indicate a higher mean in the control group (e.g., increased renal disease in the control group for HF-3), while positive imbalances (denoted in blue) indicate a higher mean in the treated group (e.g., higher proportion of treated patients were admitted to hospitals with cardiac bypass surgery capabilities).

While multiple imbalances were found for each measure, the nature of the imbalances varied by measure. Among the HF measures, imbalances in facility characteristics were common and indicated that the resources and type of hospital may influence whether patients received these measured processes of care. Several pre-existing conditions were much more common in the population not treated versus the population treated with angiotensin converting enzyme inhibitors or angiotensin receptor blockers as measured under HF-3, including renal disease (which could make physicians less likely to prescribe these drugs), anemia, and chronic obstructive pulmonary disease. For the SCIP measures, the research team found notable imbalances in procedure types, highlighting variation in process rates across procedure types. In particular, the group differences suggest that patients who underwent colon surgery were less likely to receive measured SCIP processes, while patients with knee replacement were more likely to receive these measured care processes than patients undergoing another procedure. The greatest numbers of imbalances in pre-existing conditions were observed for HF-2, *SCIP-Inf-9: Urinary Catheter Removed on Postoperative Day 1 (POD 1) or Postoperative Day 2 (POD 2) With Day of Surgery Being Day Zero* (NQF #0453) (SCIP-Inf-9), SCIP-VTE-1, SCIP-VTE-2, OP-4, and OP-6. Imbalances were generally consistent across specification periods.¹⁸

The research team applied propensity score methods to minimize the imbalances between treatment groups shown in Figure 8-4. Figure 8-5 shows an example of this process for the set of facility-level characteristics for the HF-2 measure. In these plots, the white region indicates values of SMD that suggest covariate imbalance, and the shaded region values indicate covariate balance. Without weighting, much higher mean bed counts and higher percentages of hospitals with a CABG facility, cardiac unit, catheterization laboratory, and teaching status are evident at hospitals where HF-2 patients were treated, as well as lower regional mortality rates (orange lines). Both weighting (red line) and matching (blue line) greatly reduced these differences, bringing nearly all mean differences across measures and specification periods close to zero.

Figure 8-4: Summary of Observed Covariate Imbalance

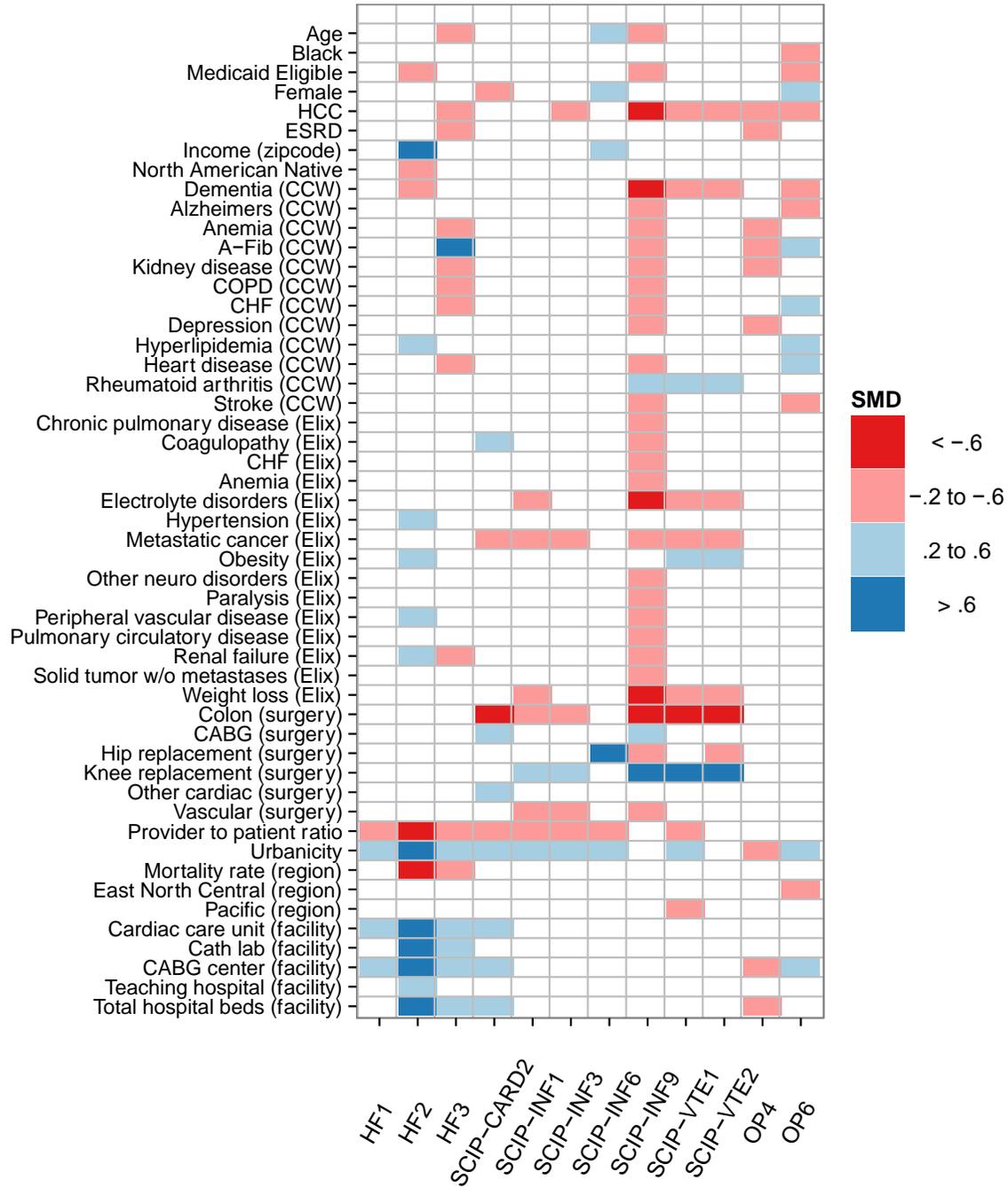
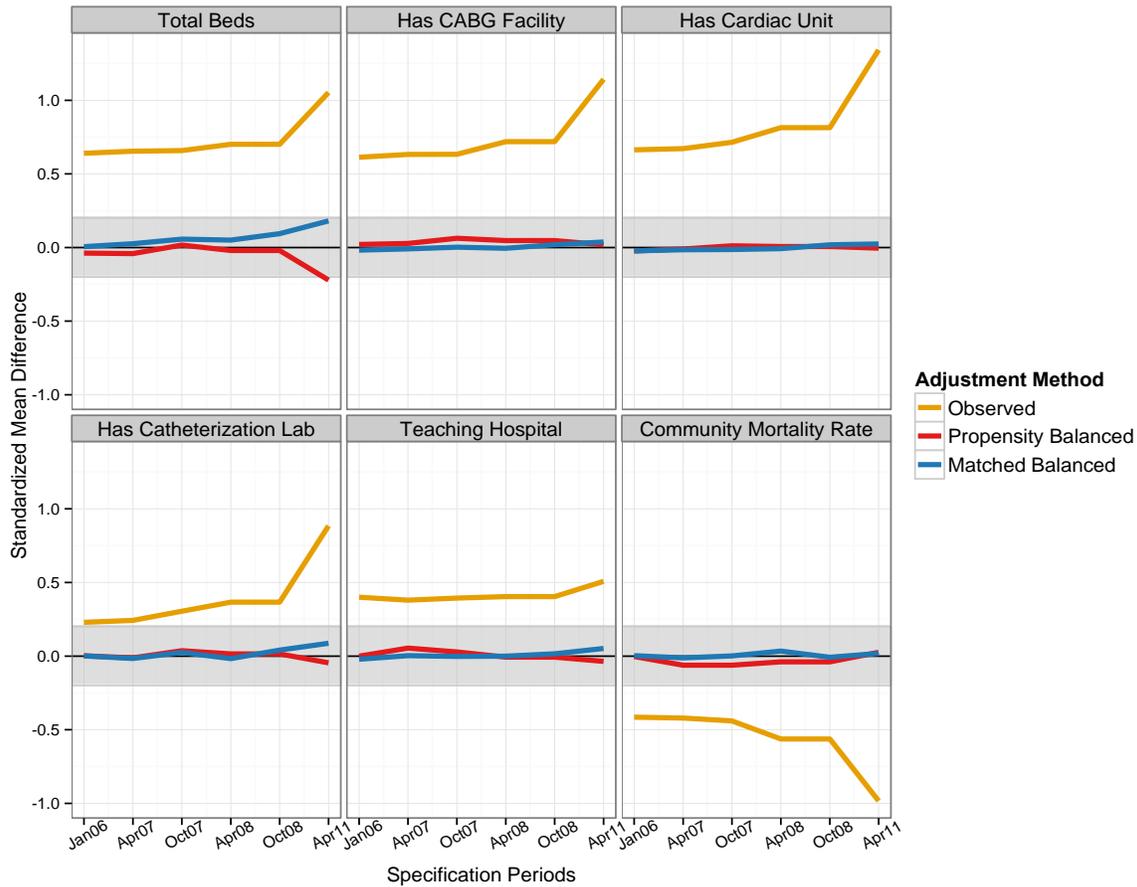


Figure 8-5: Covariate Balance With and Without Adjustment for Selected Covariates for the HF-2 Measure



The propensity scores developed for each measure and period achieved balance in mean differences for all covariates, with few exceptions. In the most recent specification periods for the SCIP-VTE-1 and SCIP-VTE-2 measures, severe imbalances in the representation of procedure type were reduced but not removed with propensity score weighting.¹⁸ As an additional safeguard against remaining imbalances, the research team adjusted the effect estimates for covariates with a weighted SMD of magnitude 0.05 or more, a so-called doubly-robust estimate.

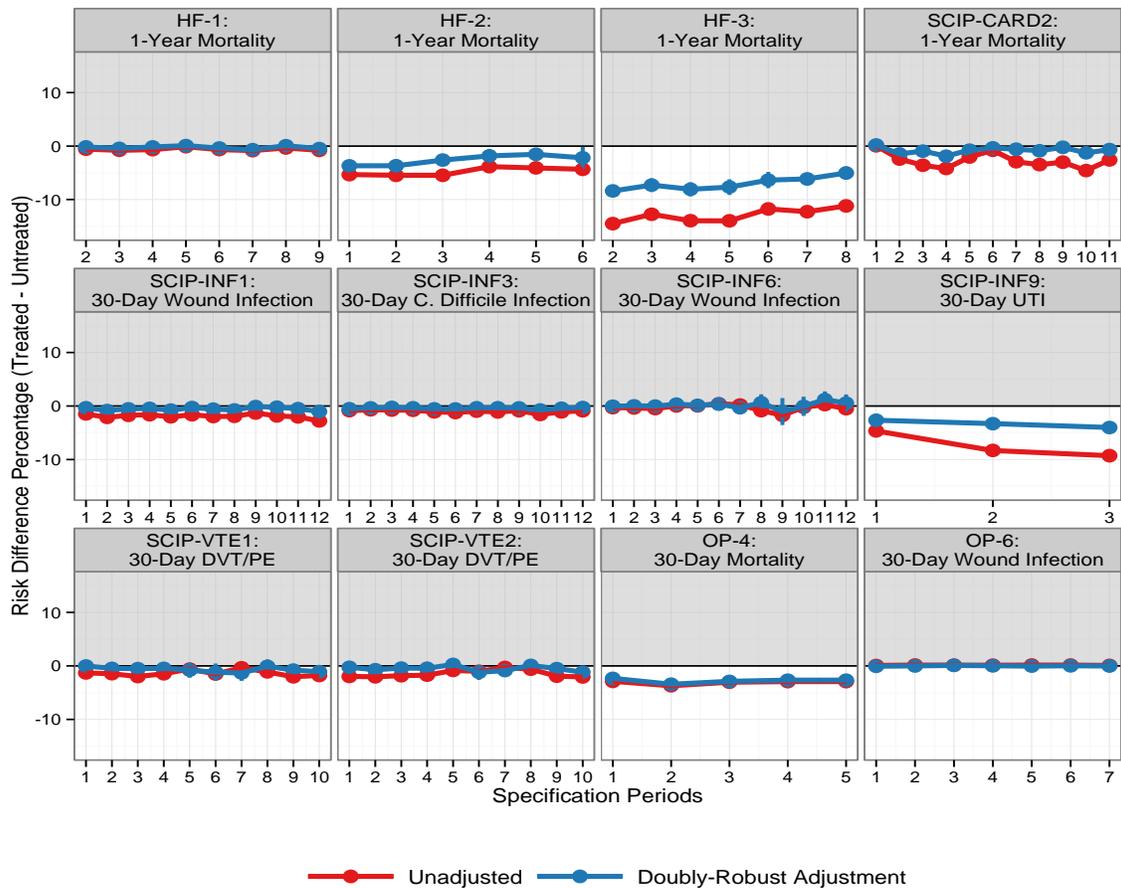
The ratio of providers to the 65-and-over population in the county where care occurred was much less variable in the treatment population than the population that was not treated for most of the prioritized process measures. This means that extremely low or high provider-to-population ratios were more likely among patients who were not treated than patients who were treated. Propensity-score weighting improved the similarity in variance between treatment groups except for the provider-to-population ratio on the HF-2 measure.

As a sensitivity analysis, the research team compared the balance achieved with inverse propensity score weighting to a 1:2 matched sample. The blue line in Figure 8-5 for the HF-2 population provides an example of the matched sample balance. The research team found that matching achieved similar balance in means and somewhat improved balance in the variances of continuous measures.

Objective 1—Determine the Association of Changes in Performance Rates for Clinical Process-of-Care Measures and Changes in Patient Outcomes

After adjustment for covariate imbalances, the research team found that adherence to process-of-care measures was generally associated with statistically significant reductions in mortality (Figure 8-6). HF-1, HF-2, and HF-3 were all associated with a reduction in one-year all-cause mortality. The effect was small for HF-1 (-0.4 percent overall, Table 8-5); moderate for HF-2 (-3.4 percent overall); and large for HF-3 (-8.5 percent overall). Significant reductions in targeted infection rates were associated with the SCIP-Inf-1, SCIP-Inf-3, and SCIP-Inf-9 processes of care (Figure 8-6). SCIP-Inf-9 adherence was associated with large reductions in lower urinary tract infection at 30 days (-4.4 percent overall, Table 8-5), while the risk reductions for 30-day wound infection associated with SCIP-Inf-1 and SCIP-Inf-3 were less than 1 percent. The research team did not find SCIP-Inf-6 adherence to be associated with reduced wound infection rates. SCIP-VTE-1 was associated with a 0.8 percent reduction in 30-day DVT or PE, while SCIP-VTE-2 was associated with 0.7 percent reduction (Table 8-5). OP-4 was associated with a moderate 2.9 percent absolute risk reduction of 30-day all-cause mortality. The research team found no significant association between OP-6 and 30-day wound infection rates.

Figure 8-6: Outcome Differences Between Patients Who Received and Did Not Receive Process Measure



There was less consistency in the effects of process measures on readmission (Table 8-5). The risk of HF readmission within one year was significantly greater for patients who received HF-1 (hazard ratio [HR]=1.04 overall) and HF-2 (HR=1.13 overall) processes than those who did not, while the risk of readmission for patients who received the HF-3 process was significantly less (HR=0.92). SCIP-Card-2 had no significance effect on all-cause readmission within one year overall (HR=0.98).

When the research team members compared the adjusted process effects between specification periods, they found significant heterogeneity for all the process measures (Q-test, p value <0.01). In particular, all the Hospital IQR Program measures had at least one process-outcome pair with an I^2 greater than 65 percent, and the between-period standard deviation was greater than 0.5 percent for the majority of the risk difference estimates and greater than 0.01 for the hazard ratio estimates (Table 8-5). This heterogeneity indicates that there were differences in the estimated effects of process by specification period. This could signal differences in the effect of process among excluded populations that have changed as measure specifications have been revised. Due to the large samples included in certain specification periods, heterogeneity measures will be sensitive to differences in effects that may be too small to be clinically relevant. For this reason, the magnitude of the standard deviation in effect by period may be a more useful measure of the impact of these changes than the Q-test of heterogeneity.

Table 8-5: Summary of Adjusted Risk Difference Effects

Measure	Outcome	Summary of Process Effects, (95% Confidence Interval)	Between Specification Period Standard Deviation
		Absolute Risk Differences, With Process Versus Without Process	
HF-1	1-Year Mortality	-0.4 (-0.5, -0.3)	0.2
HF-2	1-Year Mortality	-3.4 (-4.0, -2.7)	1.5
HF-3	1-Year Mortality	-8.5 (-9.7, -7.2)	3.3
SCIP-Inf-1	30-Day Wound Infection	-0.9 (-1.1, -0.7)	0.8
SCIP-Inf-3	30-Day <i>C. difficile</i> Infection	-0.6 (-0.7, -0.5)	0.4
SCIP-Inf-6	30-Day Wound Infection	-0.0 (-0.2, 0.2)	0.5
SCIP-Inf-9	30-Day UTI	-4.4 (-6.0, -3.0)	2.8
SCIP-Card-2	1-Year Mortality	-1.3 (-1.7, -0.9)	1.3
SCIP-VTE-1	30-Day DVT/PE	-0.8 (-1.0, -0.5)	0.8
SCIP-VTE-2	30-Day DVT/PE	-0.7 (-1.0, -0.5)	0.8
OP-4	30-Day Mortality	-2.9 (-3.0, -2.7)	0.2
OP-6	30-Day Wound Infection	0.0 (0.0, 0.1)	0.1
		Hazard Ratios, With Process Relative to Without Process	
HF-1	1-Year CHF Readmission	1.04 (1.02, 1.05)	0.02
HF-2	1-Year CHF Readmission	1.13 (1.09, 1.16)	0.04
HF-3	1-Year CHF Readmission	0.92 (0.90, 0.93)	0.02
SCIP-Card-2	1-Year All-Cause Readmission	0.99 (0.98, 1.00)	0.01

Objective 2—Assess the Impact of Changes in Performance Rates for Clinical Process-of-Care Measures on the Health Outcomes of the Target Medicare Population

For each process-outcome association indicating a statistically significant reduction in clinical outcomes for patients who received the process measure, the research team estimated the total reduction in adverse clinical outcomes that might have been achieved in the Medicare population

eligible for the measure given the adjusted estimates of process effect and the observed increase in process delivery between 2006 and 2012. Since CMS began implementing Hospital IQR Program and Hospital OQR Program, increases in process rates have resulted in hundreds of thousands of newly treated patients. In other words, although the research team cannot directly link such changes to public reporting, it is possible to calculate the expected number of patients who would not have received the process if process rates had stayed at their levels when CMS measurement began (Table 8-6). The largest gains in newly treated patients were for SCIP-Inf-3 (more than 400,000); and HF-1, SCIP-Inf-1, and SCIP-Inf-6 (each more than 200,000).

For each process associated with a statistically significant reduction in clinical outcomes for patients receiving the measured process of care, the research team estimated the total reduction in adverse clinical outcomes that might have been achieved due to increased adherence rates. Since 2006, increased adherence to the HF process-of-care measures was associated with 8,200 averted deaths within one year of care episodes (Table 8-6) with HF-3 accounting for more than half the reduction. Together, increased delivery of SCIP-Inf-1, SCIP-Inf-3, SCIP-Inf-6, and SCIP-Inf-9 was associated with a reduction of more than 5,000 infection events within 30 days of the care episode, with SCIP-Inf-1 and SCIP-Inf-3 accounting for 80 percent of the reduction. Greater delivery of SCIP-Card-2 was associated with a reduction in 2,000 deaths within one year of the care episode. Increases in SCIP-VTE-1 and SCIP-VTE-2 were associated with the reduction of 1,100 DVT or PE events within 30 days of care. Because of the population overlap for these measures, their independent population impact could not be determined. In Hospital OQR Program, increased delivery of OP-4 had a more modest effect on its targeted outcome than the IQR measures, having an estimated association of 163 fewer deaths within 30 days of the outpatient encounter.

Table 8-6: Estimates of Decreased Outcomes With Increased Process Delivery, 2006–2012^{cxvi}

Measure ^{cxvii}	Outcome	Total Newly Treated, 2006–2012	Total Reduction in Adverse Clinical Outcomes Associated With Process-of-Care, 2006–2012 (95% CI)
HF-1	1-Year Mortality	279,282	1,050 (761, 1,338)
HF-2	1-Year Mortality	49,609	1,664 (1,361, 1,967)
HF-3	1-Year Mortality	66,124	5,598 (4,780, 6,408)
SCIP-Inf-1	30-Day Wound Infection	229,204	2,034 (1,513, 2,555)
SCIP-Inf-3	30-Day <i>C. difficile</i> Infection	421,296	2,393 (1,956, 2,830)
SCIP-Inf-6	30-Day Wound Infection	235,894	71 (-355, 497)
SCIP-Inf-9	30-Day UTI	33,114	1,451 (924, 1,980)
SCIP-Card-2	1-Year Mortality	136,400	1,793 (1,235, 2,351)
SCIP-VTE-1	30-Day DVT/PE	155,688	1,142 (750, 1,536)
OP-4	30-Day Mortality	5,729	163 (155, 171)

^{cxvi} SCIP-Inf-9 was not introduced until 1/2010; the period of evaluation included for SCIP-VTE-1 and SCIP-VTE-2 began on 1/2007; the period of evaluation included for OP-4 began on 4/2008.

^{cxvii} Due to the overlap in the HF-2 and HF-3 populations, the HF-2 newly treated excludes patients newly treated with the HF-3 process; similarly, due to the near complete overlap in the SCIP-VTE-1 and SCIP-VTE-2 populations, only the combined impact is reported.

Table 8-7 shows the process measures and their association with clinically meaningful effects, the current endorsement status by the National Quality Forum (NQF), and the measure status in CMS quality measurement programs. Several of the measures evaluated in this study have subsequently been removed from the Hospital IQR and OQR Programs, including six of the eight measures that did not have an association with clinically meaningful effects. The four measures found to be associated with clinically meaningful effects (as defined in Table 8-5) remain endorsed by NQF.

Table 8-7: Process Measures and NQF Endorsement

Measure	Program	NQF Status ^{cxviii}	CMS Status as of 2017 Payment Determination
Measures found to be associated with clinically meaningful effects			
HF-2	Hospital IQR Program	Endorsed – Reserve ^{cxix,39}	Removed for FY 2017 payment determination ⁴⁰
HF-3	Hospital IQR Program	Endorsed	Removed for FY 2016 payment determination ⁴¹
SCIP-Inf-9	Hospital IQR Program	Endorsed	Retained as a voluntary electronic health record-reported measure ⁴²
OP-4	Hospital OQR Program	Endorsed	Active (CY 2017) ⁴³
Measures not found to be associated with clinically meaningful effects			
HF-1	Hospital IQR Program	De-endorsed	Removed for FY 2016 payment determination ⁴⁴
SCIP-Inf-1	Hospital IQR Program	Endorsed	Retained as a voluntary electronic health record-reported measure ⁴²
SCIP-Inf-3	Hospital IQR Program	Endorsed	Removed for FY 2017 payment determination ⁴⁰
SCIP-Inf-6	Hospital IQR Program	Endorsed – Reserve	Removed for FY 2017 payment determination ⁴⁰
SCIP-Card-2	Hospital IQR Program	Endorsed	Removed for FY 2017 payment determination ⁴⁰
SCIP-VTE-1	Hospital IQR Program	De-endorsed	Removed for FY 2015 payment determination ³²
SCIP-VTE-2	Hospital IQR Program	Endorsed	Removed for FY 2017 payment determination ⁴⁰
OP-6	Hospital OQR Program	De-endorsed	Removed for CY 2017 payment determination ⁴³

Secondary Analyses

Models that accounted for heterogeneity in base rates and process rates among hospitals did not change the findings above.¹⁸ However, there was notable between-hospital variation in process effects. The magnitude of the variation suggests that as many as 15 percent of hospitals showed no positive process-outcome association for most measures with a significant association overall. This inconsistency in process effects suggests that these analyses did not fully account for important differences in hospital and patient characteristics.

Using matched analyses, the research team examined the average process effect in the control population. These analyses showed risk difference effects that were comparable to the adjusted treatment effects (ATEs) of the main analysis. Based on this observation, the research team would expect patients who did not receive care to have a benefit similar to that observed for those who were treated.

An additional analysis of the effect of HF-1 for the shorter-term outcome of 30-day mortality or CHF readmission found no significant process-outcome association.

^{cxviii} NQF endorsement status verified August 15, 2014.

^{cxix} NQF reserve status is a mechanism to retain endorsement of reliable and valid quality performance measures that have overall high levels of performance and little variability.

Sensitivity Analyses

The research team identified no substantial differences in the process-outcome associations when performing analyses using simple weighting with inverse propensity scores, multivariable regression adjustment with no weighting, and adjustment with 1:2 matching. The consistency of the findings across approaches suggests that these results were robust and were not unduly influenced by the method of adjustment used.¹⁸

Because patients may be eligible for multiple process-of-care measures, especially those targeting the same population (e.g., HF patients, patients undergoing surgery), all or part of the effects the research team observed for individual measures may be *overlap effects*. HF-2 and HF-3 exemplify the potential for overlap effects, because systolic dysfunction must be identified before it can be treated. To better separate the effects of HF-2 and HF-3, the research team compared the effect of HF-2 adherence on one-year mortality for beneficiaries eligible for both treatments to those eligible for HF-2. Similar effects were found in each subgroup. The research team saw no evidence that the benefit of HF-2 can be attributed solely to the effect of HF-3, but this does not exclude the possibility that treatments or other interventions could influence the overall -3.4 percent mortality reduction associated with HF-2.

C-statistics were used to evaluate how well the included covariates controlled for differences in risk between treatment groups (a c-statistic=0.5 indicates the lowest possible control and 1.0 is the best possible). The c-statistics ranged from 0.67 to 0.85, indicating adequate explanatory power for all measures and excellent explanatory power for certain measures. The HF measures had the lowest c-statistics, indicating that unexplained variation in outcomes was greatest for these measures. The expansion of the models to include a number of utilization and health status measures derived from CMS-Part B physician claims reduced risk difference effects by less than 0.3 percentage points in general,¹⁸ and did not change qualitative conclusions about the clinical or statistical significance of the effects in the main analysis.

The research team conducted additional analyses for the HF-3 measure and one-year mortality outcome—the process-outcome pair with the largest adjusted effect. As shown in Figure 8-4, the research team found that persons in the HF-3 measure population who did not receive an ACE inhibitor or ARB were 50 percent more likely to have a diagnosis of renal failure at the time of the encounter than patients who received treatment. This raises the concern that the Elixhauser indicator of renal failure may not capture important differences in severity of disease that also lead to differences in treatment. For example, physicians may avoid providing ACE inhibitors or ARBs to beneficiaries with severe renal dysfunction, who likely have worse outcomes independent of not receiving the process measure. Not accounting for this instance of selection bias may lead analysts to incorrectly estimate that providing the process intervention leads to better outcomes. However, the research team found that HF-3 adherence was associated with similar reductions in one-year mortality in patients with and without renal disease.

Renal disease was consistently the characteristic that differed the most between the HF-3 treated group and non-treated group, suggesting a possible selection bias. However, in subgroup analyses by renal disease status, there was no evidence that selection bias affected the main results for the HF-3 measure.

The number of imbalances in risk adjusters the research team found for the process measures in this study raises the concern that other omitted risk adjusters could affect findings on patients who did and did not receive a given process of care. While a propensity-score method adjusts for imbalances in observed characteristics, it does not adjust or balance omitted (unknown) factors. An omitted variable could bias the process-outcome relationships reported in this study if it (1) has a greater representation in one of the treatment groups, and (2) is associated with the targeted outcome. When the difference in the representation of an omitted confounder is large, its influence on the health outcome may not have to be large to substantially bias the process-outcome effect. Because large imbalances were common for the covariates observed in this study, it is possible that moderate influences on effect substantially reduced the observed process-outcome effects. However, it is unlikely that an omitted variable could explain risk difference effects as large as those found for HF-3.¹⁸

Discussion

Based on a patient-level analysis that included more than 3.9 million inpatient encounters and nearly 200,000 outpatient encounters between 2006 and 2012, the research team found significant evidence that increased adherence to process-of-care measures was associated with improved health outcomes. In analyses of 10 measures from the Hospital IQR Program (HF-1, HF-2, HF-3, SCIP-Inf-1, SCIP-Inf-3, SCIP-Inf-6, SCIP-Inf-9, SCIP-Card-2, SCIP-VTE-1, and SCIP-VTE-2) and two measures from the Hospital OQR Program (OP-4 and OP-6) that adjusted for imbalances between the treatment and non-treatment comparison groups, the research team found clinically and statistically significant absolute reductions in all-cause mortality rates associated with the HF-2, HF-3, SCIP-Card-2, and OP-4 process measures. SCIP-Inf-9 was the only surgical measure assessed that was associated with a large absolute reduction in 30-day infection rates. The research team combined these findings with the observed increases in process adherence rates during 2006–2012 and estimated that HF measures may have been associated with 8,200 fewer deaths within one year of care. Increased SCIP-Inf adherence may have been associated with 5,000 fewer infections within 30 days of the initial surgeries, and increases in SCIP-VTE-1 and SCIP-VTE-2 adherence may have been associated with 1,100 fewer DVT or PE events within 30 days of the initial surgeries.

Prior research on the link between HF process measures and outcomes has generally found mixed mortality benefits. When Werner and Bradlow compared one-year mortality rates of hospitals in the highest quartile of performance on HF measures to those in the lowest quartile during 2004, they found a non-significant absolute risk difference of 2.2 percent.⁶ Jha et al. found that the mortality outcomes of HF patients treated in top-performing hospitals in 2005 had a significant 0.4 percent absolute reduction in inpatient mortality.¹² In an analysis of 5,791 HF patients admitted to 91 U.S. hospitals in 2003 and 2004, Fonarow et al. found a significant reduction in 60- to 90-day all-cause mortality associated with the use of ACE inhibitors in the presence of left ventricular systolic dysfunction.⁴⁵ They found no significant association between mortality rates and delivery of discharge instructions or the evaluation of left ventricular systolic dysfunction.

Studies of process-outcome relationships for SCIP measures have generally shown negative results. In an analysis of Medicare patients undergoing surgery in 2005 and 2006, Nicholas et al. (2010) found no association between a hospital's performance on Hospital Compare's surgical

measures and mortality, surgical-site infection, or VTE.⁴⁶ Using a patient-level analysis of discharges between 2006 and 2008, Stulberg and colleagues found a significant reduction in post-operative infections with improved scores on a composite of the SCIP-Inf measures but not with improved scores on any individual measures.⁴⁷

The present assessment revealed stronger evidence of benefit for several HF, SCIP-Inf, and SCIP-VTE process-of-care measures than had been previously reported. Absolute reductions in one-year mortality associated with HF-2, HF-3, and SCIP-Card-2 measures were large. SCIP-Inf-9 had a strong effect on reduced 30-day UTI, and the outpatient measure, OP-4, was associated with a nearly 3 percent absolute risk reduction in 30-day mortality. Improved performance on SCIP-Inf-1, SCIP-VTE-1, and SCIP-VTE-2 were associated with weaker but relevant clinical improvements in infection and thromboembolism rates, respectively. In addition to finding several measures associated with large clinical benefits, the research team also found measures associated with small clinical benefits. HF-1 was associated with a statistically significant but modest 0.4 percent decrease in absolute risk of one-year mortality. SCIP-Inf-3 was associated with a 0.6 percent absolute risk reduction in 30-day *C. difficile* infection. Neither SCIP-Inf-6 nor the OP-6 surgical measures were significantly associated with lower post-operative infection rates.

Readmission findings were more complex than the findings for mortality or complication rates. While the associations for HF-1 and HF-2 suggested that the relative risk of CHF readmission was higher for treated patients, HF-3 had a significantly lower readmission rate. Prior work has also reported negative associations between process and readmission. Specifically, Fonarow and colleagues reported increased readmissions for patients of the OPTIMIZE-HF registry who received HF-1 and HF-2 process interventions. An important issue in assessing the effect of process on readmission is to appropriately control for differences in survival, as patients with a higher rate of survival would necessarily have more opportunity to experience a readmission than patients with a lower rate of survival. This study made an effort to account for differences in survival when evaluating readmission outcomes using survival analysis techniques that treated death as a censoring event, independent of readmission risk. However, a competing risk model might be more appropriate than a standard survival analytic approach as it would not assume that death is unrelated to readmission risk.

A few of the study findings (especially the large mortality reductions associated with HF-3) showed substantially larger benefits of process when compared to other studies. The findings may be explained by several methodological differences between the present study and earlier work. Principally, the research team used both process and outcome data at the patient level, whereas many prior studies examined associations between hospital-level performance and patient outcomes. Although hospital-level analyses are important to address many policy questions, hospital-level associations may not reflect patient-level associations, a situation that is sometimes referred to as the ecological fallacy.⁴⁸ Discrepancies between hospital- and patient-level associations are expected to be even greater when there are fewer differences in hospital performance rates, as is the case now for many monitored processes of care.

Several of the measures evaluated in this study have subsequently been removed from the Hospital IQR Program (Table 8-7). The present status of process measures was not a consideration in the study measure selection criteria because the objective of the work was to

evaluate the historical impact of processes monitored by CMS between 2006 and 2012. However, five of the eight measures that did not appear to have had a meaningful impact on population health in this study have been or are scheduled to be removed. This finding provides support for concerns raised by CMS about the continued benefit of monitoring these measures and the actions taken.

Decisions to maintain NQF endorsement and continued use of a process measure in a CMS program could be informed by evidence about the relationship between the implemented process measure and targeted health outcomes. Other factors that can inform these decisions include, but are not limited to, clinical trial evidence, whether the measure has topped out, and whether measure use has produced other negative effects. If a process measure is being developed, the measure developer should define the outcome(s) linked to the care process and provide an estimate of how much improvement in the outcome might be gained for each increment of improvement in process compliance, and on which data to base these estimates. This process will facilitate the ability to conduct effectiveness studies of the process outcome relationship when the measure is used in practice. A measure user should ensure that data are systematically collected on the outcome identified by the measure developer to assess the impact of the process in practice. Because process-outcome evaluations are typically based on observational data, the measure user can strive to minimize the limitations inherent to observational studies and carefully consider these limitations when interpreting the findings.

Conclusions

Key Findings

A positive relationship was found between performance on a limited number of CMS measures and positive patient outcomes; however, the effects were variable, and a small number of process measures were estimated to have an impact on the health of the eligible Medicare population. Specific findings include:

- ◆ For measures that were positively associated with their targeted outcome, improved performance rates in the eligible Medicare population between 2006 and 2012 were associated with between 7,000 and 10,000 lives saved through improved performance on inpatient hospital heart failure process measures, and between 4,000 and 7,000 infections averted through improvement in performance on inpatient hospital surgical process measures.
- ◆ Strong associations between the receipt of a care process and improved patient health outcomes were identified for the following measures: HF-2, HF-3, SCIP-Inf-9, SCIP-Card-2, and OP-4.
- ◆ Modest associations were found for the timing of prophylactic antibiotics (SCIP-Inf-1 and SCIP-Inf-3) and perioperative VTE prophylaxis (SCIP-VTE-1 and SCIP-VTE-2).
- ◆ Three of the 12 evaluated measures, HF-1, SCIP-Inf-6, and OP-6, had either a small or no association with their targeted health outcome. Important differences were noted in the magnitude of the population health impact associated with improved performance among measures within the same measure set (e.g., heart failure, surgical care infections). For example, HF-3 was associated with many more averted deaths than HF-1 or HF-2; also, SCIP-Inf-3 was associated with more averted complications than SCIP-Inf-9.

Actions to Consider

The measures evaluated in this chapter were studied because they have a strong clinical evidence base, identifiable targeted outcomes, and sufficient data to support a controlled analysis of the association between the process and the targeted outcome. Measures meeting these criteria were generally found to be positively associated with improved patient health. However, the absence of a measureable outcome in CMS data, as well as insufficient data to control for differences in case mix between those who do and do not receive the process of care, precluded the ability to assess the impact on health outcomes for a majority of process measures currently monitored by CMS.

Action to Consider: Assess the adequacy of current data systems for studying process-outcome relationships. Efforts to evaluate the health impact of CMS use of clinical process measures are currently hindered by a lack of sufficient data on targeted outcomes and control variables, i.e., variables that describe differences between patients who do and do not receive the clinical process of care, needed for meaningful studies of process-of-care effectiveness. Important steps to improve future evaluations of program impact include assessing current data systems to determine whether the necessary data elements on clinical performance, targeted outcomes, and control variables are present, and, where data are missing, consider how these elements can be collected to assess the impact associated with the use of process measures.

Action to Consider: Quantify and monitor process-outcome linkages during measure implementation. Within the CMS reporting programs, measure developers should quantify the strength of the relationship between the process of care and the corresponding health outcome and monitor this linkage during implementation. The measure developers could quantify this linkage as part of the annual and three-year comprehensive review required by NQF.

Action to Consider: Develop more outcome measures, including patient-reported outcomes, since process measures are not uniformly achieving better health outcomes. Prioritization of outcome measures and the development of standard methodologies (e.g., risk adjustment) may facilitate outcome measure development.

Reference List

- (1) Fung CH, Lim YW, Mattke S, Damberg C, Shekelle PG. Systematic review: the evidence that publishing patient care performance data improves quality of care. *Ann Intern Med.* 2008;148(2):111-123.
- (2) U.S. Department of Health and Human Services. Totten AM, Wagner J, Tiwari A, O'Haire C, Griffin J, Walker M, investigators. Public Reporting as a Quality Improvement Strategy—Closing the Quality Gap: Revisiting the State of the Science. Rockville, MD: Agency for Healthcare Research and Quality: Evidence Report No. 208. (Prepared by the Oregon Evidence-based Practice Center under Contract No. 290-2007-10057-I.). Evid Rep Technol Assess (Full Rep 2012); AHRQ Pub. No. 12-E011-EF(208.5):1-645.
- (3) Silverman SL. From randomized controlled trials to observational studies. *Am J Med.* 2009;122(2):114-120.
- (4) Ryan AM. Effects of the Premier Hospital Quality Incentive Demonstration on Medicare patient mortality and cost. *Health Serv Res.* 2009;44(3):821-842.
- (5) Werner RM, Bradlow ET. Relationship between Medicare's hospital compare performance measures and mortality rates. *JAMA.* 2006;296(22):2694-2702.
- (6) Werner RM, Bradlow ET. Public reporting on hospital process improvements is linked to better patient outcomes. *Health Aff (Millwood).* 2010;29(7):1319-1324.
- (7) Damberg CL, Sorbero ME, Lovejoy SL, Martsolf G, Raaen L, Mandel D. *Measuring Success in Health Care Value-Based Purchasing Programs: Summary and Recommendations.* 2014.
- (8) Sidorenkov G, Haaijer-Ruskamp FM, de Zeeuw D, Bilo H, Denig P. Review: relation between quality-of-care indicators for diabetes and patient outcomes: a systematic literature review. *Med Care Res Rev.* 2011;68(3):263-289.
- (9) Chatterjee P, Joynt KE. Do cardiology quality measures actually improve patient outcomes? *J Am Heart Assoc.* 2014;3(1):e000404.
- (10) Bradley EH, Herrin J, Elbel B, et al. Hospital quality for acute myocardial infarction: correlation among process measures and relationship with short-term mortality. *JAMA.* 2006;296(1):72-78.
- (11) Freemantle N, Cleland J, Young P, Mason J, Harrison J. Beta blockade after myocardial infarction: systematic review and meta regression analysis. *BMJ.* 1999;318(7200):1730-1737.
- (12) Jha AK, Orav EJ, Li Z, Epstein AM. The inverse relationship between mortality rates and performance in the Hospital Quality Alliance measures. *Health Aff (Millwood).* 2007;26(4):1104-1110.
- (13) Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S. Does pay-for-performance improve the quality of health care? *Ann Intern Med.* 2006;145(4):265-272.

- (14) Setoguchi S, Warner Stevenson L, Stewart GC, et al. Influence of healthy candidate bias in assessing clinical effectiveness for implantable cardioverter-defibrillators: cohort study of older patients with heart failure. *BMJ*. 2014;348:g2866.
- (15) Ryan AM, Burgess JF Jr, Tompkins CP, Wallack SS. The relationship between Medicare's process of care quality measures and mortality. *Inquiry*. 2009;46(3):274-290.
- (16) Dunkelgrun M, Boersma E, Schouten O, et al. Bisoprolol and fluvastatin for the reduction of perioperative cardiac mortality and myocardial infarction in intermediate-risk patients undergoing noncardiovascular surgery: a randomized controlled trial (DECREASE-IV). *Ann Surg*. 2009;249(6):921-926.
- (17) McGory ML, Maggard MA, Ko CY. A meta-analysis of perioperative beta blockade: what is the actual risk reduction? *Surgery*. 2005;138(2):171-179.
- (18) Kovalchik S, Shetty K, Ganz D et al. National Assessment of Process Performance and Patient Outcomes Technical Report. Santa Monica, CA: RAND Corporation; 2014. Report No.: WR-1056.
- (19) Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long Term Care Hospital Prospective Payment System and Fiscal Year 2015 Rates; Final Rule. *Fed Regist*. 2014;79(163):49854-50449.
- (20) Van Walraven C, Austin PC, Jennings A, Quan H, Forster AJ. A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. *Med Care*. 2009;47(6):626-633.
- (21) Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care*. 1998;36(1):8-27.
- (22) Li P, Kim MM, Doshi JA. Comparison of the performance of the CMS Hierarchical Condition Category (CMS-HCC) risk adjuster with the Charlson and Elixhauser comorbidity measures in predicting mortality. *BMC Health Serv Res*. 2010;10:245.
- (23) Wennberg DE, Sharp SM, Bevan G, Skinner JS, Gottlieb DJ, Wennberg JE. A population health approach to reducing observational intensity bias in health risk adjustment: cross sectional analysis of insurance claims. *BMJ*. 2014;348:g2392.
- (24) Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11(5):550-560.
- (25) Stuart EA. Matching methods for causal inference: A review and a look forward. *Stat Sci* 2010;25(1):1-21.
- (26) Imai K, Ratkovic M. Covariate balancing propensity score. *J R Statist Soc B*. 2014;76(1):243-263.
- (27) Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Stat Med*. 2004;23(19):2937-2960.

- (28) Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557-560.
- (29) Vanderweele TJ, Arah OA. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology*. 2011;22(1):42-52.
- (30) Lawson EH, Louie R, Zingmond DS, et al. A comparison of clinical registry versus administrative claims data for reporting of 30-day surgical complications. *Ann Surg*. 2012;256(6):973-981.
- (31) QualityNet. Hospital inpatient quality reporting (IQR) program measures (Calendar Year 2014 discharges). Available at: <https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier3&cid=1138900298473>. Accessed 8/20/2014.
- (32) Medicare Program; Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long Term Care; Hospital Prospective Payment System and Fiscal Year 2014 Rates; Quality Reporting Requirements for Specific Providers; Hospital Conditions of Participation; Payment Policies Related to Patient Status; Final Rule. *Fed Regist*. 2013;78(160):50782.
- (33) Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long Term Care Hospital Prospective Payment System and Fiscal Year 2015 Rates Final Rule. *Fed Regist* 2014;79(163):49854-50449.
- (34) Medicare Program; Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long-Term Care Hospital Prospective Payment System and Fiscal Year 2013 Rates; Hospitals' Resident Caps for Graduate Medical Education Payment Purposes; Quality Reporting Requirements for Specific Providers and for Ambulatory Surgical Centers. *Fed Regist*. 2012;77(170):53503-53521.
- (35) Centers for Medicare & Medicaid Services, The Joint Commission. Specifications Manual for National Hospital Inpatient Quality Measures. Report No.: 3.0c. Available at: <http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1228695698425>. Accessed 11/18/2014.
- (36) Centers for Medicare & Medicaid Services, The Joint Commission. Specifications Manual for National Hospital Quality Measures. Report No.: 2.2b. Available at: <http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1163010419895>. Accessed 11/18/2014.
- (37) Centers for Medicare & Medicaid Services, The Joint Commission. Specifications Manual for National Hospital Inpatient Quality Measures. Report No.: 2.5b. Available at: <http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1203781887871>. Accessed 11/18/2014.
- (38) Centers for Medicare & Medicaid Services, The Joint Commission. Specifications Manual for National Hospital Inpatient Quality Measures. Report No.: 3.3b. Available at:

<http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1228760129036>. Accessed 11/18/2014.

- (39) National Quality Forum. Measures endorsed with “Reserve Status.” Available at: http://www.qualityforum.org/About_NQF/CSAC/docs/Reserve_Status_Memo.aspx. Accessed 11/20/2014.
- (40) Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long Term Care Hospital Prospective Payment System and Fiscal Year 2015 Rates; Final Rule. *Fed Regist.* 2014;79(163):50208.
- (41) Medicare Program; Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long Term Care; Hospital Prospective Payment System and Fiscal Year 2014 Rates; Quality Reporting Requirements for Specific Providers; Hospital Conditions of Participation; Payment Policies Related to Patient Status; Final Rule. *Fed Regist.* 2013;78(160):50779-50783.
- (42) Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long Term Care Hospital Prospective Payment System and Fiscal Year 2015 Rates; Final Rule. *Fed Regist.* 2014;79(163):50201-50346.
- (43) Medicare and Medicaid Programs: Hospital Outpatient Prospective Payment and Ambulatory Surgical Center Payment Systems and Quality Reporting Programs; Physician-Owned Hospitals: Data Sources for Expansion Exception; Physician Certification of Inpatient Hospital Services; Medicare Advantage Organizations and Part D Sponsors: CMS-Identified Overpayments Associated with Submitted Payment Data; Final Rule. *Fed Regist.* 2014;79(217):66940-66944.
- (44) Medicare Program; Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long Term Care; Hospital Prospective Payment System and Fiscal Year 2014 Rates; Quality Reporting Requirements for Specific Providers; Hospital Conditions of Participation; Payment Policies Related to Patient Status; Final Rule. *Fed Regist.* 2013;78(160):50680-50783.
- (45) Fonarow GC, Pan W, Saver JL et al. Comparison of 30-day mortality models for profiling hospital performance in acute ischemic stroke with vs without adjustment for stroke severity. *JAMA.* 2012;308(3):257-264.
- (46) Nicholas LH, Osborne NH, Birkmeyer JD, Dimick JB. Hospital process compliance and surgical outcomes in Medicare beneficiaries. *Arch Surg.* 2010;145(10):999-1004.
- (47) Stulberg JJ, Delaney CP, Neuhauser DV, Aron DC, Fu P, Koroukian SM. Adherence to surgical care improvement project measures and the association with postoperative infections. *JAMA.* 2010;303(24):2479-2485.
- (48) Piantadosi S, Byar D, Green S. The ecological fallacy. *Am J Epidemiol.* 1988;127(5):893-904.

Chapter 9—Measure Relationships: Patient-Reported Hospital Experiences and Predicted Medicare Costs



Question on Effectiveness

Is there a relationship between hospital-level patient-reported experience of care and risk-adjusted Medicare spending for hospitalization and 30-day post-hospital care?

Abstract

Background: The Centers for Medicare & Medicaid Services promotes the aims of better care, better health, and lower costs. This study examines the relationship between two of these goals—better care and lower costs. Literature suggests that a focus on patient-centered care, as measured by instruments such as the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) Survey, may be associated with improved health outcomes and lower costs of care. This study examines the relationship between HCAHPS scores and Medicare costs resulting from utilization of services from three days pre-admission to hospitalization and 30 days post-discharge as measured by the risk-adjusted Medicare Spending per Beneficiary (MSPB) ratios.

Methods: This study included hospitals with data for both MSPB ratios and HCAHPS scores during 2012 (N = 3,147). The research team used linear regression analysis to examine the extent to which variation in MSPB ratios, generated from all Medicare fee-for-service discharges, could be predicted by selected HCAHPS scale scores, generated from all adult hospital discharges, both before and after controlling for important variables found in the literature. These literature-identified characteristics included the geographic region; the average age and age distribution of patients; the size of the hospital; the female, Black, and Hispanic representation among the hospitals' patients; designation as a Safety Net Hospital; and designation as a Rural Hospital. The HCAHPS scales, selected for the model due to their relatively low correlations with each other and relatively high correlation with MSPB ratios, were:

- ◆ “Percent of patients who reported that they ‘Always’ received help as soon as they wanted.”
- ◆ “Percent of patients who reported that their doctors ‘Always’ communicated well.”
- ◆ “Percent of patients who reported that staff ‘Always’ explained about medicines before giving it to them.”

Results: A small, statistically significant, and methodologically persistent relationship was found between positive patient-reported inpatient experiences and lower-than-expected pre-admissions, hospitalization, and 30-day post-discharge costs. Hospitals where patients reported higher satisfaction with their inpatient experiences as measured by HCAHPS had somewhat lower-than-expected Medicare fee-for-service costs associated with admissions, as measured by MSPB ratios compared with hospitals with less positive patient-reported experiences of care ($R^2 = 0.06$, $r = 0.24$).

Conclusions: The available evidence suggests that patient-experience reports as measured by HCAHPS may reflect important elements that contribute to hospital efficiency (lower cost with higher quality). Additional analyses at the individual patient level are needed to confirm the findings at the facility level with the publicly reported data and to ensure a congruent match between the populations represented by HCAHPS scores and hospital-related costs. Qualitative inquiry could be used to gather variables of interest for facilities, independent of patients.

Background

The strategic goals for the Centers for Medicare & Medicaid Services (CMS) for 2013–2017 include providing better care, better health, and lower costs of care for all Americans.¹ As a country with high healthcare costs but worse-than-expected health outcomes relative to many parts of the world,² the challenge is to identify the best, most efficient means by which to improve care, while ensuring care remains patient-centered and of optimal quality for everyone.

This study seeks to explore the relationship between patient experience and cost. CMS, hospitals, healthcare systems, health plans, and other providers use a variety of metrics to inform discussions on lowering cost, improving quality, increasing positive patient-reported experiences of care, and improving other priority areas. Often, these metrics are associated with financial incentives to encourage the desired change. CMS has implemented incentives for measures through programs such as the Hospital Value-Based Purchasing Program (Hospital VBP Program), through which hospitals receive a financial incentive for high and/or improved performance on a set of measures.

Concerns have been expressed, however, that as healthcare providers are given incentives to reduce costs of care or increase positive patient-reported experiences of care, quality will suffer. Lowering costs by reducing important aspects of care, such as nurse staffing, or increasing positive patient-reported experiences of care by prescribing expensive drugs desired by patients are two potential unintended consequences.^{3,4} Nonetheless, measures that are patient-centric and

meet the needs of multiple stakeholders in the healthcare system may encourage the desired outcome of high quality care at lower cost. Measures that address the needs and goals of these multiple stakeholders promote “shared value”⁵ among patients, hospitals, and the nation’s healthcare system in pursuit of high quality care at lower cost.

Patient advocates have long asserted that patient experience is an important component of understanding the quality of healthcare. The literature shows that patient experience is theoretically and empirically related to quality, patient experience provides a unique understanding of quality not captured by clinical measures, and patients are good judges of the clinical quality of their care.⁶⁻⁹ Instruments used to capture patient satisfaction in these studies include Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) and the Press Ganey patient satisfaction survey. Other authors have cautioned, however, that clinical quality and patient experience are distinct domains, and patient experience should not be used as a proxy for quality.⁸ To resolve the somewhat conflicting results, further research needs to be done to determine if patient experience is a valid hospital quality indicator.⁴

The National Quality Forum (NQF), the national consensus-based entity that endorses quality healthcare measures, believes that efficiency (encompassing both quality and cost) is best understood when cost and resource use measures are linked with measures of care quality, including those of patient-reported outcomes.¹⁰ Patient-reported outcomes include several domains, one of which is patient experience.¹¹ This study is intended to conduct an exploratory examination of the relationship between the patient-reported outcome domain of patient experience and cost.

Although the literature is limited concerning direct justification for a relationship between patient experience and cost, an indirect case is made through evidence of the relationships of patient experience and clinical processes and outcomes. Doyle et al. found evidence for a positive relationship between patient experience and increased patient safety and clinical effectiveness across several conditions and medical settings.¹² Furthermore, 30-day hospital readmissions and mortality, two important outcome measures, have also been linked to patient-reported experiences.¹³⁻¹⁵ Recently, these relationships were found in a study using a large national database of measures of surgical quality and patient experience, linking positive patient experiences to higher performance on process measures, lower readmission rates, lower mortality rates, and shorter lengths of stay.¹⁶

Costs of care have a tenuous relationship with quality in the literature. Baicker and Chandra have found that states with higher Medicare spending have lower-quality care.¹⁷ Cooper found that states with higher healthcare spending have better-quality care.¹⁸ Conversely, regional studies have found no relationship between higher-spending geographic regions and quality of care.^{19, 20} These mixed findings are frequently explained as the result of other unmeasured factors or differences in methodologies across studies.^{18, 21, 22} Peikes et al., citing a 2008 Medicare Payment Advisory Commission (MedPAC) report,²² noted that hospital admissions and readmissions are key drivers of Medicare costs.

Two additional mechanisms have been suggested in the literature that might explain why a relationship between patient experience and cost should exist. First, positive patient experience may contribute to lowering costs, perhaps through greater patient engagement²³ and ability to

conduct self-care after treatment,²⁴ leading to fewer readmissions. Second, lower costs may be reflective of an element not captured by clinical process measures or patient-reported experiences of care, such as greater commitment by hospital management to improving processes and reducing costs.²⁵ Given these conflicting findings, detailed and large-scale investigations of the relationship between patient-reported experiences of care and cost of care are needed to better understand the apparent contradictions.

To examine this relationship, the research team employed two measures used by CMS to examine patient experience and costs—HCAHPS (NQF #0166) and Medicare Spending per Beneficiary (MSPB),^{cxx} respectively. HCAHPS is a national survey of patients’ perspectives of their hospital experience, publicly reported on the *Hospital Compare* website at the hospital level. HCAHPS was selected because it is a national survey of patient experience used in the Hospital IQR Program and Hospital VBP Program.²⁶ HCAHPS has been endorsed by NQF since 2005 and underwent a rigorous scientific review process prior to endorsement and implementation.²⁷ As previously described, evidence has suggested an indirect link between improved patient experience and reduced costs through improved quality, although the literature is not consistent on the topic. HCAHPS data can be used to compare hospitals on local, regional, and national levels and include issues such as communication with doctors and nurses, responsiveness of staff, pain management, conveyance of discharge information and medicines, cleanliness and quietness of patients’ rooms, whether patients would refer family and friends to the hospital, and an omnibus rating of the hospital. A portion of the total performance score that is used to determine a hospital’s payment for the Hospital VBP Program is linked to the HCAHPS survey. Thus, hospitals have a financial incentive to perform well on the HCAHPS survey.

The other measure, MSPB, is a ratio of the risk adjusted, price standardized Medicare costs to that of the national median. This measure also includes trimming of certain populations and limiting extreme values in the data to reduce the effect of outliers.^{cxxi} This measure is calculated from Medicare fee-for-service (FFS) claims. The measure captures unexpected Medicare expenses or savings in a ratio format, rather than actual spending in dollars. MSPB is used in public reporting for the IPPS hospitals.²⁸ CMS currently plans to include MSPB in the Hospital VBP Program beginning in Fiscal Year 2015.

Through a better understanding of the relationships between patient experiences of care and costs that were not anticipated for that care, the results from this research question might inform the discussion as to whether “incurred” costs or “saved” costs were reflected by differences in patient-reported experiences of that care. Although costs can be addressed in various ways, including cost of encounters, lengths of stay, or readmission rates, MSPB was chosen because it is an NQF-endorsed measure used by CMS for public reporting. Further, MSPB is in use in the Hospital VBP Program and is the most reliable and valid source of the relevant information publicly available to address the objective of this study.

^{cxx} NQF-endorsed title: *Payment-Standardized Medicare Spending Per Beneficiary (MSPB)* (NQF #2158).

^{cxxi} The MSPB Measure Information Form located at the following link provides more information about the methodology used in the development of the MSPB measure:

<http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=OnetPublic%2FPage%2FQnetTier4&cid=1228772057350>

Objective

The objective of this study is to examine the relationship between hospital-level patient-reported experience of care and risk-adjusted Medicare spending for hospitalization and 30-day post-hospital care.

Methods

Measures and Data Sources

Multiple regression analysis and dichotomized cross-tabulation validation models are presented to examine the relationship between patient experience and differences in MSPB ratios. The data required to address the full model that include the influential covariates found in the literature are available from the Hospital IQR Program. The research team accessed the Hospital IQR Program data from the CMS Abstraction and Reporting Tool (CART, 2012) and the CMS data warehouse (2011). Additional data included the American Hospital Association (AHA) Survey data purchased from the AHA (2011), MSPB, and HCAHPS data downloaded from *Hospital Compare* for data collected from January 1, 2012 through December 31, 2012. A hyperlink to a list of HCAHPS and MSPB measures used in this study is provided in Appendix i-4. The list includes the NQF endorsement status, NQF number, and both the measure title used by the CMS program and the measure title used by NQF. Table 9-1 lists the CMS programs used for the current analysis.

Table 9-1: Programs Included in Analysis

Setting	Program	Abbreviation
 Hospital	Hospital Inpatient Quality Reporting Program	Hospital IQR Program
	Hospital Value-Based Purchasing Program	Hospital VBP Program

Analysis

Objective—Examine the Relationship Between Hospital-Level Patient-Reported Experience of Care and Risk-Adjusted Medicare Spending for Hospitalization and 30-Day Post-Hospital Care

The relationship between the MSPB ratios modeled by HCAHPS responses is examined by adjusting for hospital characteristics found in the literature that are known to be associated with either hospital inpatient costs or patient-reported experiences of care. Potential covariates identified in the literature are presented in Appendix 9-1, with their shortened literature citations.

To examine the quantitative relationship between patient-reported experiences of hospital care and MSPB ratios, separately reported HCAHPS subscales were used for two reasons:

- ◆ Separately reported subscales were used rather than aggregates because separate results are topic-specific and are potentially actionable by being influenced through policy and provider changes.
- ◆ The literature on HCAHPS has shown that the subscales capture distinct, although related, domains (shown by the correlation matrix in Appendix 9-2) delineating specific areas for opportunities for improvement.

The hospital-level^{cxixii} correlations across the 10 HCAHPS subscale scores are frequently moderate to high, with one bivariate correlation exceeding 0.90 (Appendix 9-2). This situation can cause partial collinearity problems in a multiple regression model. High correlations among explanatory variables may cause interpretational paradoxes and should be avoided whenever possible while constructing a policy-relevant model. The process for choosing the HCAHPS subscale items to use in modeling differences in MSPB ratios included four steps:

1. Encompass the widest number of HCAHPS domains without over-fitting the model or encountering analytically inconsistent partial collinearity issues.
2. Start with the HCAHPS subscale that has the strongest bivariate correlation with MSPB.

“Percent of patients who reported that they ‘Always’ received help as soon as they wanted,” $r = -0.23$ with MSPB (the negative correlation means higher HCAHPS scores are associated with relative savings compared with expected spending).

3. Choose a subscale that balances a relatively low correlation (the numerical correlations are moderate to high) with the first HCAHPS variable selected and a relatively high correlation with MSPB.

“Percent of patients who reported that their doctors ‘Always’ communicated well,” $r = 0.68$ with first HCAHPS subscale and $r = -0.21$ with MSPB.

4. Choose a third HCAHPS subscale with a balance of relatively low correlations with the first and second HCAHPS variables selected and a relatively high correlation with MSPB.

“Percent of patients who reported that staff ‘Always’ explained about medicines before giving it to them,” $r = 0.75$ and $r = 0.69$, $r = -0.22$ for correlations with the first and second HCAHPS subscale and MSPB, respectively (the negative correlation means higher HCAHPS scores are associated with relative savings compared with expected spending).

The subscales assess “top-box” scores only.^{cxixiii} The items that form each of the 10 reported subscale scores are also shown in Appendix 9-2.

^{cxixii} Correlations at the patient level are somewhat lower. HCAHPS patient-level correlations can be found online at <http://www.hcahpsonline.org/SummaryAnalyses.aspx>.

^{cxixiii} “Top-box” scores include the most positive response HCAHPS categories, including “Always” for 5 HCAHPS Composites, “Yes” for Discharge Information, “9” or “10” for Hospital Rating, and “Definitely” for Recommend the Hospital. More information on calculation of HCAHPS scores can be found online at <http://www.hcahpsonline.org/Files/Calculation%20of%20HCAHPS%20Scores.pdf>.

Limitations

Related but not fully congruent variables of interest are used in this study. HCAHPS responses are collected for a random sample of all adult hospital inpatients from two to 84 days after discharge. MSPB ratios are calculated for Medicare FFS beneficiaries and extend to 30 days post discharge. Matched groups would require patient-reported experiences of care (HCAHPS results) limited to discharged Medicare FFS beneficiaries, which would be a subset of the HCAHPS results. For the purposes of the current study, it is assumed that hospital systems of care are generally applied across all populations so that similar levels of positive patient-reported experiences of care would exist for all groups.

Additionally, literature suggests that populations vary in reports of care by the nature of their culture,^{29,30} disease status,¹³ and potentially other factors. HCAHPS scores are patient-mix adjusted for patient-level characteristics, including age, service line, education, and self-reported overall health. Thus, some patient-level characteristics are already present in the model via the HCAHPS patient-mix adjustment. Unadjusted person-level data would make it possible to match beneficiaries' reports of care and MSPB data and allow for an examination of whether reports of care might vary by other aspects of the patient population not captured by hospital-level data.

Perhaps most importantly, the MSPB ratio includes expenses associated with 30 days post-discharge, including readmissions. An assumption can be made that hospital admissions followed by readmissions result in higher costs and are less satisfying to patients than outcomes that do not include readmissions. Controlling for readmissions could substantially change the coefficients for HCAHPS in the model as an unknown percentage of HCAHPS respondents are readmitted to the hospital. Further, doing so (as well as earlier-than-planned discharges) could also remove an important aspect of what is being measured as patient-reported experiences of care from the HCAHPS coefficients. Although the model might provide a better fit, the coefficients for the HCAHPS subscales would change their meaning and would need interpretive modifications, such as "... controlling for readmissions," and thereby remove their effect on the relationship between cost and patient-reported experiences of care.

Results

The three selected HCAHPS subscale scores together explain 6.00 percent of the variability (i.e., $R^2 = 0.06$) in MSPB ratios. The explanatory power of each of the three selected HCAHPS subscales is small, with standardized coefficients ranging from -0.06 to -0.11, having had bivariate correlations ranging from -0.21 to -0.23. The total explained variance in MSPB ratios increases approximately threefold (i.e., from 0.06 to 0.22) in the full model using the three selected HCAHPS subscales and the literature-identified covariates. The sizes of the standardized coefficients for the HCAHPS subscale scores change somewhat, but none of the values reached +/- 0.20, indicating degrees of association. As a frame of reference, -0.21 was the average geographic region effect found in the model. By comparison, none of the HCAHPS effects on MSPB was as strong as the average geographic region effect.

Table 9-2: Model 1: MSPB Modeled by Selected HCAHPS Subscales and Literature-Identified Covariates

Dependent Variable = MSPB Ratios	Std Beta	Zero-Order Corr	p value	VIF ^{cxiv}
HCAHPS Scales Only				
Percent of patients who reported that they "Always" received help as soon as they wanted	-0.11	-0.23	<0.01	2.63
Percent of patients who reported that their doctors "Always" communicated well	-0.06	-0.21	0.02	2.18
Percent of patients who reported that staff "Always" explained about medicines before giving it to them	-0.10	-0.22	0.01	2.73
HCAHPS Scales With Literature-Identified Covariates				
Percent of patients who reported that they "Always" received help as soon as they wanted	-0.07	-0.23	0.01	3.17
Percent of patients who reported that their doctors "Always" communicated well	-0.16	-0.21	<0.01	2.79
Percent of patients who reported that staff "Always" explained about medicines before giving it to them	-0.08	-0.22	<0.01	2.85

As indicated by the zero-order correlations among the HCAHPS subscale scores and the correlations with the covariates in the model, it is not surprising that each of their coefficients substantially changes in magnitude, both among themselves and again when the covariates are added, including their relative strengths and contributions to MSPB ratio modeling. Yet, the direct effect of each subscale of the HCAHPS points to three conclusions. First, satisfied patients cost less than expected, adjusting for the literature-identified covariates. Second, the magnitudes of the relationships are relatively small when compared against geographic region effects that were already partially removed through the creation of the MSPB ratio and further removed in the present model. Third, the relationship between better patient-reported experiences of care and lower costs is persistent across each of the examinations shown (i.e., bivariate, restricted model, and full model).

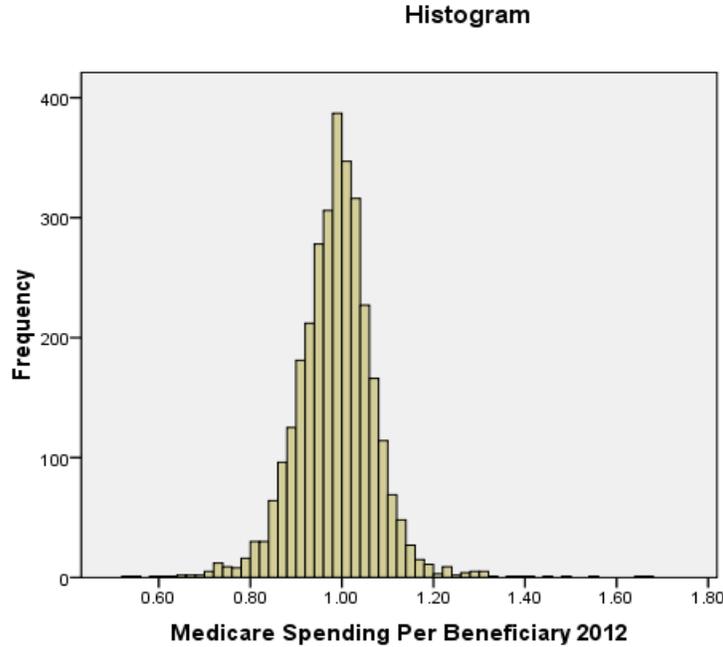
To place the size of the effect in perspective, a 10-percentage point difference in the HCAHPS subscale score for “Percent of patients who reported that their doctors ‘Always’ communicated well,” which was the strongest (i.e., most influential) HCAHPS variable for explaining differences in MSPB ratios, is associated with an average of a 0.03 shift in the ratio. To put the 0.03 value in perspective, approximately three-quarters of all MSPB scores are within one standard deviation of the mean, or within 0.09 MSPB ratio points from the mean score of 0.98. In essence, large changes in HCAHPS subscale scores are associated with relatively small changes in expected hospital and 30-day follow-up expenses. Yet, the finding is robust to the degree of sophistication for the model, as previously discussed.

Finding higher explanatory ability between unexpected costs (or savings) and HCAHPS subscale scores and covariates than were found herein would have been unexpected. Although the range of MSPB ratios in the model is from 0.53 to 1.66, the mean of 0.98 with its standard deviation of 0.09 results in a coefficient of variation of 0.09, which is exceedingly narrow. This narrow range

^{cxiv} VIF is the Variance Inflation Factor, often used for collinearity diagnostic purposes when building regression models.

of values for MSPB is shown in Figure 9-1. Again, MSPB provided the best available data to use in the model for the current research question.

Figure 9-1: Distribution of MSPB Ratios for 3,147 Hospitals



As explained, Figure 9-1 suggests that the model could have relatively weak predictive validity. A common method for evaluating the fit of a model (i.e., a measure of predictive validity) is through the cross-tabulation of actual grouped values by predicted grouped values. For the current model, values above and below 1.00 for MSPB were chosen as the cut-off point because 1.00 is the theoretical national average that should result from the MSPB creation process. Table 9-3 shows the model fit test results for the MSPB ratio and its differences across hospitals.

Table 9-3: MSPB Model Fit

Total Agreement = 70.9 Percent		Actual MSPB Ratio Group		Total
		Low	High	
Predicted MSPB	Low	1,544	521	2,065
	High	396	686	1,082
Total		1,940	1,207	3,147

The model correctly predicts over- or under-expected hospital and follow-up expenses 70.9 percent of the time $[(1,544 + 686) / 3,147]$, where approximately 50 percent would be expected from random guessing. When the kappa coefficient is calculated to evaluate total agreement beyond chance, the result is 0.37, which is below the pre-determined minimum of 0.50 that the literature suggests should be required to be retained as policy-relevant.³¹ For this reason, though the results from the multiple regression analyses seem favorable and interpretable, the examined fit of the model is not sufficient to support policy recommendations. The full regression model is shown in Appendix 9-3.

The model was also applied with the first principal component for the 10 reported HCAHPS scores as the dependent variable and MSPB as the explanatory variable of interest, with and without the same covariates. The results for the extent of the association between patient-reported experiences of care and unexpected costs or savings were attenuated due, likely in part, to the smaller number of predictor variables (i.e., one versus three predictor variables). These attenuated results mean that the policy relevance of the relationship is also reduced. Given the limitations to the policy relevance of the initial model, the results for this version are not shown here but are found in Table 9-3-3 and 9-3-4 within Appendix 9. The amount of explained variation for the HCAHPS Principal Component Analysis variable by the MSPB variable is half that found when modeling the MSPB variable using the three selected HCAHPS variables (i.e., R^2 is 0.03 versus 0.06, respectively).

Discussion

Hospitals with higher HCAHPS survey scores are associated with lower-than-expected average Medicare costs for the period from three days before admission through 30 days after discharge, both before and after controlling for externally influential variables found to be important from the literature. Although persistent and statistically significant, the strength of the relationship was not sufficiently large to warrant policy recommendations.

Nonetheless, the high degree of statistical significance between lower-than-expected costs and greater positive patient reports of care suggests that further work should be conducted in this area to better understand the relationship. For example, readmissions could be driving the result for three reasons. First, a patient whose outcome is not as favorable and who requires being readmitted may feel less positive about his or her initial care and would incur greater costs than a patient with a typical hospital outcome. Second, a patient who fares better than average would probably feel better about his or her care and would not require as many follow-up services as an average patient and would be less likely to be readmitted, which would represent lower costs. Third, the timing of patient responses to HCAHPS may be an important issue for understanding the relationship between patient-reported experiences of care and unexpected costs or savings. In the current study, a patient may have responded to the survey during a period that exceeds that encompassed by the MSPB. Hospitals administer the HCAHPS between two and 84 days after discharge. A patient could have responded to the HCAHPS survey before developing a complication such as a post-operative infection and a need for a readmission within the 30-day post-discharge period included in the MSPB. Had the patient responded to the HCAHPS survey after the readmission, the patient's survey results might be substantially different, while the MSPB variable would capture the readmission costs in both cases.

In general, patient-reported experiences, as measured by HCAHPS, may reflect important elements that contribute to hospital efficiency (lower cost/higher quality); however, an analysis using person-level data is required before patient-level explanatory characteristics can be separated from provider characteristics in hierarchical models. An unknown percentage of patients who receive the HCAHPS survey are readmitted to the hospital; however, it may be possible to develop hospital-level readmission estimates for HCAHPS respondents using administrative data in future analyses. Additionally, models that are informed by qualitative surveys of subject matter experts might select important facility and patient characteristics associated with high quality care at reduced costs that have yet to be identified in the literature.

Future directions entail conducting hierarchical analyses to account for patient-level characteristics independently from facility characteristics using information gained from qualitative surveys, in addition to the literature.

Conclusions

Key Findings

Hospitals with discharged inpatients reporting greater positive experiences as measured by HCAHPS scores had lower-than-expected costs associated with admissions and 30-day post-discharge care, as measured by MSPB ratios. The relationship was statistically weak (i.e., standardized beta coefficients less than 0.20) but was persistent. This persistence was evidenced by the model retaining statistical significance in the presence of the literature-identified covariates, which included the geographic region of the hospital; the mean age of the patients; the standard deviation of the age; the total beds; the percentages of Black, Hispanic, and female patients; and a designation as a Safety Net Hospital and/or Rural Hospital.

The available evidence suggests that patient-reported experiences, as measured by HCAHPS, may identify important characteristics of patient care quality that are also associated with lower hospital and related costs.

Actions to Consider

CMS is implementing strategies to transform the healthcare delivery system, which will ensure patients receive better care at a lower cost. Although the existing literature shows that patient experience is theoretically and empirically related to quality, the literature is limited regarding direct justification for a relationship between patient experience and cost. This study aims to examine the relationship between patient experience and cost; however, policy relevance of the results is limited by the differences in the populations between all adult hospital discharges and Medicare FFS discharges and by the inability to separate facility from patient characteristics. If patient-level data are available for HCAHPS, MSPB, and readmissions, then hierarchical models that use both facility and patient data separately could model the characteristics of each group, independent of the effects of the other groups to improve the policy relevance of the model. These hierarchical models using literature-identified facility and patient characteristics could be informed by a qualitative study of characteristics thought by subject matter experts to be associated with high quality performance at reduced patient costs. Additional possibilities for further analyses include using the full range of HCAHPS values rather than only the “top-box” scores.

Action to Consider: Align eligible patient populations for cost and resource use measures with select quality measure eligible populations to facilitate further research regarding the relationship between quality and cost.

Action to Consider: Conduct a qualitative and quantitative study to understand the characteristics of an organization or provider that are associated with high performance on outcome and cost measures. Explore what factors may account for positive patient outcomes in facilities with lower-than-expected costs; for example, improved care coordination may improve patient experience, while contributing to lower cost.

Reference List

- (1) Centers for Medicare & Medicaid Services. *CMS Strategy: The Road Forward, 2013-2017*. Baltimore, MD: Centers for Medicare & Medicaid Services; 2014.
- (2) Murray CJL, Frenk J. Ranking 37th—measuring the performance of the US health care system. *N Engl J Med*. 2010;362(2):98-99.
- (3) Crow R, Gage H, Hampson S, et al. The measurement of satisfaction with healthcare: implications for practice from a systematic review of the literature. *Health Technol Assess*. 2002;6(32):1-244.
- (4) Schoenfelder T. Patient satisfaction: A valid indicator for the quality of primary care? *Primary Health Care: Open Access*. 2012;2(4):1.
- (5) Porter ME, Kramer MR. The big idea: creating shared value. *Harv Bus Rev*. 2011; January-February (Reprint):1-17.
- (6) Glickman SW, Boulding W, Manary M, et al. Patient satisfaction and its relationship with clinical quality and inpatient mortality in acute myocardial infarction. *Circ Cardiovasc Qual Outcomes*. 2010;3(2):188-195.
- (7) Isaac T, Zaslavsky AM, Cleary PD, Landon BE. The relationship between patients' perception of care and measures of hospital quality and safety. *Health Serv Res*. 2010;45(4):1024–1040.
- (8) Farley H, Enguidanos ER, Coletti CM, et al. Patient satisfaction surveys and quality of care: an information paper. *Ann Emerg Med*. 2014.
- (9) Shwartz M, Cohen AB, Restuccia JD, et al. How well can we identify the high-performing hospital? *Med Care Res Rev*. 2011;68(3):290–310.
- (10) National Quality Forum. *Endorsing Cost and Resource Use Measures*. Washington DC: National Quality Forum; January 31, 2014.
- (11) National Quality Forum. *Patient Reported Outcomes (PROs) in Performance Measurement*. Washington DC: National Quality Forum; January 10, 2013.
- (12) Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. *BMJ Open*. 2013;3(1).
- (13) Boulding W, Glickman SW, Manary MP, Schulman KA, Staelin R. Relationship between patient satisfaction with inpatient care and hospital readmission within 30 days. *Am J Manag Care*. 2011;2011/02/26(1):41-48.
- (14) Epstein AM, Jha AK, Orav EJ. The relationship between hospital admission rates and rehospitalizations. *N Engl J Med*. 2011;365(24):2287-2295.

- (15) Glickman SW, Boulding W, Manary M, et al. Patient satisfaction and its relationship with clinical quality and inpatient mortality in acute myocardial infarction. *Circ Cardiovasc Qual Outcomes*. 2010;3(2):188-195.
- (16) Tsai TC, Orav EJ, Jha AK. Patient satisfaction and quality of surgical care in US hospitals. *Ann Surg*. May 30, 2014.
- (17) Baicker K, Chandra A. Medicare spending, the physician workforce, and beneficiaries' quality of care. *Health Aff (Millwood)*. 2004;Suppl Web Exclusives:w4-w97.
- (18) Cooper RA. States with more health care spending have better-quality health care: lessons about Medicare. *Health Aff (Millwood)*. 2009;28(1):w103-w115.
- (19) Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder EL. The implications of regional variations in Medicare spending, part 1: the content, quality, and accessibility of care. *Ann Intern Med*. 2003;138(4):273.
- (20) Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder EL. The implications of regional variations in Medicare spending, part 2: health outcomes and satisfaction with care. *Ann Intern Med*. 2003;138(4):288-298.
- (21) Skinner J, Chandra A, Goodman D, Fisher ES. The elusive connection between health care spending and quality. *Health Aff (Millwood)*. 2009;28(1):w119-w123.
- (22) Peikes, D, Chen, A, Schore, J, Brown, R. Effects of care coordination on hospitalization, quality of care, and health care expenditures among Medicare beneficiaries: 15 randomized trials. *JAMA*. 2009;301(6):603-618.
- (23) Hibbard JH, Greene J, Sofaer S, Firminger K, Hirsh J. An experiment shows that a well-designed report on costs and quality can help consumers choose high-value health care. *Health Aff (Millwood)*. 2012;31(3):560-568.
- (24) Rathert C, Wyrwich MD, Boren SA. Patient-centered care and outcomes: a systematic review of the literature. *Med Care Res Rev*. 2013;70(4):351-379.
- (25) Jha AK, Orav EJ, Dobson A, Book RA, Epstein AM. Measuring efficiency: the association of hospital costs and quality of care. *Health Aff (Millwood)*. 2009;28(3):897-906.
- (26) Medicare Program; Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long Term Care; Hospital Prospective Payment System and Fiscal Year 2014 Rates; Quality Reporting Requirements for Specific Providers; Hospital Conditions of Participation; Payment Policies Related to Patient Status; Final Rule. *Fed Regist*. 2013;78(160):50782.
- (27) Centers for Medicare & Medicaid Services. HCAHPS Fact Sheet (CAHPS Hospital Survey). August 1, 2013; Available at: http://www.hcahps.org/files/August_2013_HCAHPS_Fact_Sheet3.pdf. Accessed August 29, 2014.

- (28) Medicare Program; Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long-Term Care Hospital Prospective Payment System and Fiscal Year 2013 Rates; Hospitals' Resident Caps for Graduate Medical Education Payment Purposes; Quality Reporting Requirements for Specific Providers and for Ambulatory Surgical Centers. *Fed Regist.* 2012;77(170):53503-53521.
- (29) Martino SC, Weinick RM, Kanouse DE, et al. Reporting CAHPS and HEDIS data by race/ethnicity for Medicare beneficiaries. *Health Serv Res.* 2013;48(2 Pt 1):417-434.
- (30) Elliott MN, Lehrman WG, Goldstein E, Hambarsoomian K, Beckett MK, Giordano LA. Do hospitals rank differently on HCAHPS for different patient subgroups? *Med Care Res Rev.* 2010;67(1):56-73.
- (31) Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics.* 1977;33(1):159-174.

Chapter 10—Future Directions

Introduction

The 2015 Impact Report represents the most comprehensive assessment performed to date of the CMS quality measurement programs. This report includes a setting-wide approach that distinguishes it from focused, single-setting measure set evaluations conducted by measure developers and from program evaluations that examine particular program components. The findings suggest that quality measures used in CMS quality reporting programs have contributed to improving the national healthcare system in each of the three aims of Better Care, Healthy People/Healthy Communities, and Affordable Care. The transformation is occurring across the healthcare system in part due to key CMS initiatives, such as health information technology, aligned payment incentives, and quality measurement.

As new measurement programs are established, new clinical quality measures are added, and additional longitudinal data become available, a clear opportunity exists to quantify further the impact of these quality-reporting programs on patients and providers. To that end, current projections for the 2018 Impact Report include 11 reporting programs and nearly double the number of quality measures included in the 2015 Impact Report. In addition to a more comprehensive database from which to conduct trend analyses, the 2018 Impact Report as proposed will address the following topics, which are discussed in more detail in the next section:

- ◆ Quality Measure Impact on the CMS Quality Strategy Goals.
- ◆ Impact of Quality Measurement on Providers.
- ◆ Relationship between Quality and Cost.
- ◆ Impact of Quality Measurement on Patient Outcomes.
- ◆ Identification and Monitoring of Healthcare Disparities.
- ◆ Impact of Quality Measure Exclusions on Patient Outcomes.

Focal Points for Future Study

The 2018 Impact Report will highlight progress on the CMS Quality Strategy, while endeavoring to provide new insights for making informed measure and program-specific decisions. Important focal points for the 2018 Impact Report include:

Quality Measure Impact on the CMS Quality Strategy Goals

The CMS Quality Strategy goals, which are aligned with the National Quality Strategy priorities, serve as guiding principles for measure development at CMS. In the 2018 Impact Report, monitoring the proportional trends of quality measures attributed to each measure domain across reporting programs, as well as classifying the performance gap of measures within specific domains, will determine the quality measure impact on the CMS quality strategy goals.

Impact of Quality Measurement on Providers

As part of the 2015 Impact Report, the research team developed a mixed-methods approach to answer the study questions outlined below. Work completed during the 2015 Impact Report project focused on the development of provider surveys, cognitive testing of draft survey items, and preparation of Office of Management and Budget (OMB) clearance packages as mandated under the Paperwork Reduction Act (PRA). As part of the 2018 Impact Report, the research team will implement the national provider surveys. The survey approach will consist of both close-ended surveys and semi-structured qualitative interviews. The research team will field the close-ended surveys to a large, nationally representative sample of providers. The team will then conduct the semi-structured qualitative interviews with a limited number of providers to allow for in-depth examination of the impact CMS measures are having on providers. The surveys address the following study questions:

- ◆ Is the collection and reporting of quality measure results associated with changes in provider behavior?
- ◆ What factors (e.g., changes in program design, system changes) were associated with changes in performance over time?
- ◆ Are unintended consequences associated with the implementation of CMS quality measures?
- ◆ Are there barriers to providers in implementing CMS quality measures?
- ◆ What differentiates high- and low-performers?

The research team developed four surveys for providers in four settings of care that are the focus of CMS measurement programs: Hospitals, nursing homes, home health agencies, and physician offices. For the 2018 Impact Report, two surveys were selected that address the following healthcare settings and measures:

- ◆ **Hospitals:** Regarding their experiences with the Hospital Inpatient Quality Reporting Program (Hospital IQR Program) and Hospital Outpatient Quality Reporting Program (Hospital OQR Program) measures.
- ◆ **Nursing homes:** Regarding the Nursing Home Quality Initiative (NHQI) and Nursing Home Compare measures.

The OMB will review the hospital and nursing home surveys during the first year of the 2018 Impact Report project. This review will take approximately 10 to 12 months. Once approved, the hospital and nursing home surveys will be released in approximately January 2016 in preparation for the 2018 Impact Report.

Relationship Between Quality and Cost

A clearer understanding of the relationship between quality and cost is needed as CMS implements strategies to achieve the three healthcare aims of Better Care, Healthy People/Healthy Communities, and Affordable Care. The 2018 Impact Report will further explore this relationship. Available evidence suggests that patient-experience reports, as

measured by Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS), may reflect important elements that contribute to hospital efficiency (higher quality with lower cost). Qualitative and quantitative analyses are proposed to understand the characteristics of an organization or provider that are associated with high performance on outcome and cost measures. The research team will explore what factors may account for positive patient outcomes in facilities with lower-than-expected costs; for example, improved care coordination may improve patient experience and reduce readmissions, while contributing to lower costs.

Impact of Quality Measurement on Patient Outcomes

The impact of quality measures can be quantified on a number of different levels within the healthcare system, including the population, provider, and the individual beneficiary. Patient-level data acquired for the 2018 Impact Report will provide information about beneficiary characteristics that affect measure performance that may be masked by facility-level data. Related research questions could include understanding whether beneficiaries live longer, experience better outcomes, or feel better as a result of receipt of the quality of care received as reflected in performance on quality measures. To investigate this relationship, the research team proposes to analyze the impact of quality measures by focusing on health outcome and value for patients.

Identification and Monitoring of Healthcare Disparities

Trend analysis in the 2015 Impact Report indicated that widespread race and ethnicity disparities that existed at the beginning of the study period in 2006 were much less pronounced in 2012; however, disparities persist across select programs, settings, and demographic groups. For the 2018 Impact Report, the research team will analyze additional data available to characterize trends in healthcare disparities identified in the 2015 Impact Report and link these data geographically to explore patterns of disparities. Since elimination of ethnic/racial disparities is a foundational principle of the CMS Quality Strategy, the 2018 Impact Report will take a more holistic approach and expand the disparities analyses that were conducted for this report.

Impact of Quality Measure Exclusions on Patient Outcomes

Although systematic exclusions of patient populations were not identified in the 2015 Impact Report, the research team proposes analyzing the trends in exclusion rates of a subset of CMS quality measures to determine the impact of exclusions on the eligible population and the implications of population health for key high-impact conditions.

Conclusions

This 2015 Impact Report builds on knowledge of CMS measure performance trends provided in the first assessment, published in 2012,¹ and introduces a number of in-depth analyses. The trend data and analyses compiled in this 2015 Impact Report assess multiple dimensions of CMS quality measure use and results. Although certain analyses examined all measures in all 25 CMS programs, others examined selected measures in a few programs. The results for this 2015 Impact Report illustrate that CMS has achieved significant progress in improving provider performance in key aspects of care and across multiple healthcare settings.

The 2018 Impact Report will focus on nearly twice as many quality measures and programs, will explore in detail the impact of quality measurement on patient health outcomes and disparities, and will determine the impact of quality measures on the CMS Quality Strategy goals. The hospital and nursing home surveys will investigate the impact of the CMS quality measurement programs from the perspective of the healthcare provider. This periodic national impact assessment of quality measures will provide meaningful information for CMS, federal partners, and stakeholders to understand which measures have worked well and which have had less impact on quality, as measurement strategies continue to be refined. The findings gathered through this ongoing evaluation will provide information to assist CMS and all sectors of healthcare providers, payers, and communities to improve the infrastructure for quality measurement.

CMS, as a leader in healthcare transformation, uses quality measurement as a key lever to support quality initiatives that promote patient-centered care, improve transparency, and enhance value. Because quality measurement plays a pivotal role in improving quality of care, a robust quality measurement infrastructure with ongoing vigilance and evaluation is needed. A comprehensive, crosscutting assessment of the quality measures, conducted every three years (such as this 2015 Impact Report), will provide important insights for the continuous refinement and strengthening of the CMS quality measurement strategies. Each person who receives healthcare in the United States is likely to benefit from a robust quality measurement system, as healthcare professionals engage in delivery system reform to achieve better care for patients, better health for the U.S. population, and lower costs through quality improvement.

Reference List

- (1) Centers for Medicare & Medicaid Services. *National Impact Assessment of Medicare Quality Measures*. Baltimore, MD: Centers for Medicare & Medicaid Services; March 2012.